Lecture Notes in Networks and Systems 548

Mykhailo Ilchenko
Leonid Uryvsky
Larysa Globa *Editors*

# Progress in Advanced Information and Communication Technology and Systems

## Advanced Approaches to Intelligent Data Processing and Smart Networking

Springer

# Lecture Notes in Networks and Systems

Volume 548

**Series Editor**

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

**Advisory Editors**

Fernando Gomide, Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA
Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Mykhailo Ilchenko · Leonid Uryvsky ·
Larysa Globa
Editors

# Progress in Advanced Information and Communication Technology and Systems

## Advanced Approaches to Intelligent Data Processing and Smart Networking

Springer

*Editors*
Mykhailo Ilchenko
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine

Leonid Uryvsky
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine

Larysa Globa
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-16367-8 ISBN 978-3-031-16368-5 (eBook)
https://doi.org/10.1007/978-3-031-16368-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

**This volume is** a collection of the most important research results in fields of information, telecommunication and radio electronics technologies provided by different group of researchers from Ukraine in collaboration with scientists from different countries. The authors of the chapters from this collection present in-depth and extended research results in their scientific fields.

**The volume** consists of **three parts**.

**Part I Modern Challenges in Information Technologies** deals with various aspects to the analysis and solution of practically important issues of information systems in general, contains discussion about progression from big data to smart data, development of cloud-based architecture, practical implementation of Internet of Things (IoT), fundamentals of information, and analytical activities.

**Part II Modern Challenges in Telecommunication Technologies** contains original works dealing with many aspects of construction, using research and forecasting of technological and services characteristics of telecommunication systems. The presented studies of this part cover a wide range of telecommunication technologies, including wireless communication systems, multiservice transmission systems, as well as in-depth aspects of the study of methods in problems of improvement by traffic engineering.

**Part III Modern Challenges in Radio Electronics Technologies** contains actual papers, which show some effective technological solutions that can be used for the implementation of novel radio electronics systems.

For the convenience of the readers, we briefly summarize contents of the chapters accepted.

## Part I Modern Challenges in Information Technologies

The first chapter presented by the authors **I. Shmelkin, D. Matusek and A. Schill** ***Towards Role-based Context-aware Monitoring Systems*** introduces the scientific and technical principles of monitoring information technology systems during operation.

It is one of the few methods available to help administrators keep track of the monitored system's state, predict and identify errors, and assist in system repair and error avoidance. Monitoring solutions to date, however, still suffers from various trade-offs as current implementations impose architectural restrictions on monitored systems, which lead to reduced flexibility in deployment and operation. While excellent monitoring systems exist in some application domains, others are not sufficiently supported. Furthermore, most monitoring software is specialized to function with specific data formats, protocols, and data acquisition mechanisms, further reducing their flexibility. The role-based approach for modeling and implementing software promises an intuitive way of increasing the flexibility of information technology systems' modeling and implementation. Paired with technology on control loops from the domain of self-adaptive systems, we can create a reusable framework made from static role-playing building blocks that allow overcoming those limitations. In this chapter, a concept for a flexible monitoring solution on this basis is presented and discussed, which provides functioning in most application domains while minimizing constraints on the monitored system. We compare the flexibility of our concept with a collection of 15 monitoring systems based on 11 criteria. We illustrate the concept by presenting and discussing a uniform role-based model and describe its implementation afterward. Finally, we present a qualitative and quantitative evaluation of the role-based monitoring approach in comparison with two state-of-the-art monitoring frameworks in one typical use case for monitoring systems.

The chapter ***Adaptation Consistency of Distributed Role-Oriented Applications based on the Actor Model of Computation*** by **D. Matusek, T. Kluge, I. Shmelkin, T. Springer and A. Schill** focuses on the problem of structural adaptations of objects in software engineering, which can be modeled using contexts and roles. While existing programming languages support development of role-based software, approaches for distributed applications are lacking. In this paper, we present a mapping of roles to the actor model of computation, which facilitates the concurrent execution of independent objects and their communication via messing passing. A concept for strong or eventual consistent adaptation of distributed role-based applications is developed. The former approach always ensures a global consistent state of the system. The latter increases the availability of the system in the case of network congestions or failures of devices during the adaptation process by superseding a rollback of all intermediate changes. We discuss the cases for applying different consistency criteria and show the feasibility of our approach by presenting a case study of a file server application in a prototypical programming language.

The chapter presented by **L. Globa, R. Novogrudska, B. Zadoienko and Yu Junfeng** ***Ontology-Driven Approach to Scientific and Educational Information***

***Representation*** introduces an approach to ontological model development for representing information accumulated by various scientific and educational organizations. The process of such organizations functioning is associated with accumulation of a large amount of information, which in turn is used in the process of assessing the quality of their functioning. It is proposed to provide structuring and systematization of such information on the basis of ontological approach, for its further analysis and processing. The general ontology of scientific and educational organizations information identifies several interrelated components, each of which provides a description of the processes occurring in the organization, information accumulated by the organization in the process of its functioning, as well as information on quality assessment of organization. The elements of all the components of general ontology are described. The results of proposed ontology development using the ontological solutions development platform TEDAOS are given.

The chapter ***Approach to Uniform Platform Development for the Ecology Digital Environment of Ukraine*** by **L. Globa, S. Dovgyi, O. Kopiika and O. Kozlov** focuses on the problem of information system design at the national level, wherein it is necessary to ensure collecting, storing, and processing of significant amounts of heterogeneous data accumulated in the repositories of various organizations and departments, circulating in various incompatible systems and not always available for transdisciplinary analysis. To ensure full access to the information, this research proposes the approach to the Uniform Ecology Platform (UEP) development. The platform will create a single point of access to all types of heterogeneous data of the ecology digital environment of Ukraine. The platform should include business process structures, subsystem structures, information structures, and integration structures. The paper proposes five basic principles of the UEP development, namely uniform information model; common shared telecommunications infrastructure; clearly defined interfaces; independence between business processes and applied subsystems; using of a distributed system with soft links between its components. To implement these principles, the authors suggest a mathematical description of the UEP elements, tools that support the principles of its development and optimization method of successive processes between subsystems of different decomposition-based information communication systems (ICS). As an example, the paper describes global business process formation for the different ICS and their subsystems. The suggested components that are functioning together guarantee full remote access to environmental data stored in different physically distributed systems

In the chapter ***Cloud-based Technologies for Data Processing in Ukraine: International Context*** by **A. Shelestov, B. Yailymov, H. Yailymova, S. Nosok and O. Piven**, the authors discuss about big data problem in Earth observation domain. Fortunately, cloud solutions such as Amazon Web Services, Google Earth Engine, and others platforms provide an access to Sentinel-1, Sentinel-2, and Landsat data with spatial resolution from 10 to 30 m, opportunities for quick and convenient way of geospatial data processing and usage for a lot of different information products retrieval like land cover classification maps, crop state monitoring, etc. During last few years, we had experience of cloud-based technologies usage within scientific and innovation projects, which are supported by European Commission, World

Bank, United Nations Development Program, GEO Committee and have experience of open-source software development and machine (deep) learning usage in cloud environment, namely Open Data Cube. This package provides the opportunities for data collection, deployment, and provision on the base of 3D model of data representation. The developed technologies are implemented on diverse cloud platforms and solve various types of applied problems, in particular monitoring of agricultural lands, assessment of sustainable development indicators at the national level. All of these questions will be described in our chapter in more detail.

In the chapter ***The Comprehensive Approach to Big Data Preprocessing*** by **L. Globa, R. Novogrudska and M. Grebinichenko**, the authors investigated the features to optimizing the process of big data preprocessing. The existing shortcomings of input datasets that lead to decreasing of their quality in Big Data processing systems have been identified. The main methods of preprocessing of data sets are considered. The ways to Big Data clearing are described. They allow to correct distorted data. The existing approaches to development the architecture of Big Data processing systems are analyzed. The microservice architecture is proposed for their flexible processing. The possibilities of big data preprocessing have been expanded due to the improved method of data clearing based on the text data processing templates. The proposed advanced flexible complex of algorithms for big mata preprocessing with a high level of fault-tolerance allows increasing the accuracy of data further processing. Software realization (web applications) of proposed algorithms complex for data cleansing methods with proposed improvements and microservice architecture was developed. The efficiency of the proposed architecture for the big data preprocessing system based on microservices is shown on practice.

In *Mathematical Models and Informational Technologies of Crop Yield Forecasting in Cloud Environment* by **L. Shumilo, S. Drozd, N. Kussul, A. Shelestov and S. Sylantyev**, the authors presented current results in area of remote sensing data together with biophysical models. The agricultural sector plays a very important role in the country's economy. Often, crop failures are the cause of protracted financial crises, which are very difficult to overcome. Fortunately, modern scientific methods allow us to predict the yield of certain crops in selected fields based on data on soil properties. The study requires high-resolution data, the download and processing of which requires the involvement of powerful cloud infrastructures, graphics processors and other technological solutions. Involvement of these technologies and reliable forecasting of yield help to make the right decision about sowing and avoid crop failure. However, collecting data on soil properties directly from agronomists and local farmers is a very long and hard process. In addition, the data collected may be inaccurate and expensive. As a result, a process will be launched that will eventually lead to an agrarian and financial crisis. However, a solution to this problem has already been found. Remote sensing data together with biophysical models has long been used in the world.

The chapter ***Compulsoriness and Energy Efficiency in the Decentralized World of Smart Things*** by **A. Luntovskyy, T. Zobjack and M. Klymash** focuses on the energy-efficient, self-sufficient and intelligent networked nodes, and IoT devices. These are the wide range of internet-enabled household and entertainment devices

that can be centrally controlled and managed. Such devices are through medical devices that monitor the personal people condition to devices that realize monitoring of machine parks and production streets in Industry 4.0 as well as supply chain management issues. Blockchain IoT contributes to the compulsoriness, commitment in the decentralized world of smart things. Secured IoT devices can only be achieved with the combination of known crypto-technologies. Thought a little further, the step-by-step provision of different blockchain-based platforms and solutions can help to reach the desired protection goals.

In ***Standard Model of System Architecture of Enterprise IT Infrastructure*** by **S. Dovgyi and O. Kopiika**, the authors present a modern IT infrastructure model for an enterprise with own data centers and geographically dispersed organizational structure. The model includes a set of architectures, namely security, management, data storage, applications, and network. Architectures define the fundamental principles for building IT services and their relationships. Additionally, each architecture is used as the basis for setting up the requirements for the creation of IT services. The suggested principle for developing a typical system IT infrastructure is as follows: architectures define a set of services; IT services are provided to three groups of customers; IT services and customers are interrelated within five implementation scenarios; five architectures define the integration of IT services. The paper proposes to use the information technologies as IT Services, which aim at maintaining the following elements in a technically sound condition: network devices, computing equipment, data storage devices, automatic software deployment service, network service, perimeter security service, directory service, file and print service, data management service, business application service, IT management service, backup and recovery service, certificate management service, integration service. All customers of the enterprise are divided into three main categories. Where appropriate, the customers may be further grouped within each category as follows: employees, partners and partner organizations, external customers. Scenarios for the implementation of virtual offices, target groups, cabinets can be implemented within the framework of such infrastructures: own data centers, virtual department, remote office, extranet, cloud data center. A method of optimizing the process of providing for certain categories of clients with Data Center services is proposed for mathematical modeling of the corporation's IT infrastructure, which brings the solution to the incidence matrix for certain graph. As an example of designing elements of IT infrastructure, the design of optimal network architecture is considered. Examples of implementation of the proposed methodology for building IT infrastructure for service centers of PJSC "Ukrtelecom" are given.

**I. Matveev, K. Karpov, M. Iushchenko, D. Dugaev, I. Luzianin, E. Siemens and I. Chmielewski** ***(Comparative Analysis of Object Detection Methods in Computer Vision for Low-performance Computers Towards Smart Lighting Systems)*** focus on the object detection problem based on machine learning techniques, which is directed on improving both the accuracy and the detection speed of a given algorithm. Most of such approaches assume substantial amount of computational resources available to the algorithm to make it fairly efficient. Therefore, a vast majority of them are hardly feasible on low-powered and less-capable embedded

IoT devices, where the object detection tasks are equally common and usually even more challenging—considering a huge diversity in environmental conditions, camera positions and resolutions, outdoor illuminations, and deployments in both urban and offline remote environments. This paper presents a comparative research of different machine learning object detection approaches targeted specifically toward the low-performance embedded computers in a context of smart street lighting applications. The comparison includes both real and synthetic datasets preprocessed for a given detection method, conducted on a Raspberry Pi SoC platform.

In chapter ***Optimization of Control Characteristics Using the Information Model*** by **N. Lyubymova, A. Naumenko, O. Poliarus, Ye. Chepusenko and A. Lioubimova**, the authors presented the relevance and necessity of information support modernization for emission management of thermal power plants due to the change in the main technological processes and the transition to new energy sources. This revision of traditional emission control schemes and technologies is necessary to comply with natural resource management standards in Ukraine in the process of its integration into the European system. The study proposes updating environmental standards based on advanced technologies and management experience at thermal power plants in the framework of joint European pilot projects. A mathematical model of information support for the control system has been developed. It is possible to improve the quality control indicators using this information criterion: to increase the performance, reduce the manufacturer's risks, and increase the reliability by 3.5%. In the study, a statistical sample was used for monitoring flue gas pollution of a thermal power plant for 22 days with a 5-minute control.

## Part II Modern Challenges in Telecommunication Technologies

In chapter ***Resilience Improvement by Traffic Engineering Fault-Tolerant Routing in Programmable Networks*** by **O. Lemeshko, O. Yeremenko, M. Yevdokymenko, A. Mersni, V. Lemeshko and M. Persikov**, the authors proposed an approach to resilience improvement by traffic engineering fault-tolerant routing in programmable networks. The mathematical model allows formalizing software-defined network data plane construction that connects with multiple access networks. Additionally, to increase fault-tolerance, several border routers were utilized. The main aim of the solution is to improve the level of overall network resilience. At the same time, load balancing in the data plane is achieved by applying the traffic engineering concept by ensuring the packet flows distribution using primary or backup routes at the access level between several gateways that create one virtual default gateway. The technical task of fault-tolerant routing with load balancing based on a modified model has a linear programming optimization form. The model implements the protection of the default gateway, providing load balancing on the interfaces of the virtual default gateway and within the core network. The numerical research results of traffic

engineering fault-tolerant routing processes confirmed the proposed model's effectiveness in implementing the default gateway protection scheme and load balancing in the network.

**I. Strelkovskaya and R. Zolotukhin** ***(Research of Automated Control Systems Development Based on "Publish-subscribe" Technology Over Low-bandwidth Radio Networks)*** present the scientific and technical principles of the development of automated control systems (ACS) based on low-bandwidth communication networks of UHF/VHF radio stations. This development is the major problem concerning the system nodes intercommunication and information exchange. The high-throughput communication networks have solved this problem by means of algorithms and protocols based on "publish-subscribe" technology. However, it is necessary to research the ability to use such technology over low-bandwidth communication networks. This research is focused on automated control systems development based on Message Queuing Telemetry Transport (MQTT) and Data Distribution Service (DDS) by Object Management Group (OMG) protocols. The ability to use such mechanism of data exchange over governmental ACS of the low echelon management level based on VHF/UHF radio stations with high requirements to safety, liveliness, and reliability are shown. The MQTT, MQTT-SN, and DDS protocols analysis was performed. The data about service packets quantity, lost packets quantity, maximal packet size, actual volume of service, and information data was performed, and the time of communication establishment between MQTT, MQTT-SN, and DDS nodes was defined. The recommendations concerning "publish-subscribe" protocols usage in ACS development were given.

**N. Mareev, D. Kachan and E. Siemens** ***(Multipoint Data Transmission Issues in High Bandwidth-Delay Product TCP/IP Networks)*** offer new approach to analysis of modern network hardware, which nowadays is ready to provide high-speed channels across the countries and between continents, providing wide network resources to end-users. In the presence of this, algorithms of multipoint data transmission become more and more a bottleneck, utilizing available network resources not optimally, which can be a consequence of the flaws of modern software solutions and data transport protocols. In the meanwhile, several promising solutions for traffic control have been proposed in the last decade; however, the area of multipoint high-speed data transmission remains an insufficiently researched field. This work is aimed to observe the main issues of the data transmission in TCP-/IP-based wide area networks focusing mainly on congestion control algorithms and software issues in point-to-multipoint solutions. In this paper, the main issues of congestion control algorithms have been reviewed and compared in different network delays and loss probability cases. Proposed earlier Bottleneck Queue Level congestion control (BQL) algorithm has been observed in the context of decreasing the negative impact provided by widely used algorithms on the end-user side. The problem of underutilization of existing channels is described and several software methods of increasing the data throughput have been touched. Several issues of point-to-multipoint data transmission and application-layer multicast solution based on RMDT have been discussion. A final performance evaluation and comparison of TCP BBR and RMDT BQL were made with a tuned Linux TCP/IP stack. The evaluation of introduced algorithms

and their prototyping were made with Reliable Multi-Destination Transport Protocol (RMDT), a UDP-based, high-speed transport protocol. All tests have been performed in the emulated WAN environment of Future Internet Lab Anhalt (FILA).

**L. Uryvsky and K. Martynova** ***(Research of the Service Structure Influence on the Sensitivity Indicators of the Queuing System Characteristics with Priorities)*** focus on the study of the problem of access to system resources at the channel level of the OSI model. The problem of access is the objective lack of resources (time, frequency, energy, channels) for users of telecommunications services. Uneven distribution of resources creates a shortage of them among low-privileged users and a surplus for high-privileged subscribers. The problem is exacerbated by the fact that the constant development of widely available technologies and the circumstances in which society exists encourages users to increase their demands for quality of service, and as a result—a conflict of user access to telecommunications resources. The task of the study of telecommunications channels at the channel level in the framework of applied information theory is to select and further use adequate models to describe and quantify the transmission systems of information from the standpoint of access to telecommunications channels. By identifying the characteristics of the telecommunications system and the corresponding QoS, it is possible to formulate recommendations for the rational construction and organization of information transmission systems at the channel level with a known set of resources of telecommunications channels.

**I. Strelkovskaya, I. Solovskaya and Ju. Strelkovska** ***(Improving the Accuracy of User Location in the Wi-Fi Network Using Complex Spline-Functions)*** discuss approach to the analysis of factors of rapid development of various applications and services (Skyhook, Wi2Geo, Google maps, etc.), which function based on determining the current location of the user, both global GPS and local LPS. This applies, above all, to methods for determining the local location of users indoors under conditions of high concentration of users and the difficulties in radio signal distribution. In this research the usage of local methods of determining the location based on access point AP equipment within the Wi-Fi/Indoor network infrastructure was considered. A comparison of known methods for determining the local location of users was described. The research was made according different principles, namely the AOA method based on triangulation by signal level, RSS method based on RSSI signal power level measurement and TOA method based on distance trilateration to AP. It is shown that to improve the accuracy of determining the location of the user can be recommended to use a combination of several methods. In this case it is possible to compensate the disadvantages of one method by the advantages of another method. The main methods of determining the location in the Wi-Fi/Indoor network, in particular the fingerprinted method, are considered. A new method for determining the user's location based on the finite element method using complex planar splines is proposed. An example of finding the values of the user positioning error in the Wi-Fi/Indoor network is shown. The use of the proposed method will improve the accuracy of determining the coordinates of the location of the user based on the AP equipment in the Wi-Fi network and ensure the provision of LBS-based services and applications to indoor users under different conditions of their provision.

The chapter ***Principles of Building Modular Control Plane in Software-Defined Network*** by **Romanov O., Nesterenko M., Mankivskyi V. and Zhuk O.** consider some aspects related to the existing networks. These aspects deal with the networks' equipment based on monolithic physical elements. Each element of the network has features that depend on the manufacturer. This does not allow telecom operators to make changes to equipment functions, to flexibly operate resources and quickly introduce new services. The use of SDN technology can significantly increase efficiency by centralizing management and virtualizing the functions of network elements. At the same time, operators have the opportunity to independently develop applications, which significantly speeds up the process of providing new services to users. The paper shows that when implementing SDN, it is necessary to ensure compatibility with existing networks using legacy management technologies based on operations support systems (OSS). It should be possible to include a person in the control loop—an operator who will take part in solving poorly structured network management tasks that are not amenable to automation. The tasks that the SDN network management plane should solve are described. The requirements for the architecture of ONOS are formulated, which is proposed to be built in the form of a modular, symmetrical, and distributed system. A mathematical model is proposed for predicting the required SDN network resource using standard network indicators: incoming load, QoS, and throughput. The model allows solving the problem of forming a load distribution plan with given QoS indicators.

In chapter ***The Method of Using a Telecommunication Air Platform as a Flying Info-Communication Robots*** by **O. Lysenko, V. Romaniuk, A. Romaniuk, V. Novikov, V. Yavisya and I. Sushyn**, the authors are concerned with some issues related to the conducted research of effectiveness of method of improved monitoring data collection which is accumulated in the sensors of the wireless sensor network. Data collection is performed by the so-called info-communication robot under different initial conditions: network dimension, number of clusters, number of nodes in the cluster, options for constructing data collection methods, and flight strategy over nodes in the cluster. The results of efficiency comparison of using the improved method with the existing method of direct data collection and the centroid method of data collection by time criteria are given: the duration of data collection and the duration of network operation. The paper presents the analysis of four strategies of flight over cluster (only between gathering points centers; flight over critical nodes; data transfer in points that are closer to the flight route; cooperative strategy). The efficiency of the improved method of data collection from the main nodes of the clustered network was evaluated.

**S. Kravchuk and I. Kravchuk** ***(Wireless Connection of Drones to the Base Station of the Existing Terrestrial Mobile Network)*** present a novel approach to finding the location and orientation of the repeater drone antenna system in the area of the radiation pattern of the base station (BS). This approach describes development of the new three-dimensional model of such a scenario. Based on the model, an analysis of the wave propagation losses over wireless channels between a ground BS, a repeater drone, and a user terminal is carried out. A new approximation of the BS radiation pattern in the horizontal and vertical planes, the axes of which are

interconnected by the geometry of the ellipse, has been developed. The plane of such an ellipse corresponds to the cross-sectional plane of the BS directional pattern with the position of the drone, which forms a radio channel with the BS and the user terminal. The results of modeling the scenario BS-drone-terminal are presented in the form of dependences of path losses of Line of Sight (LoS) and Non-Line of Sight (NLoS) on the distances between the ground terminal, the BS, and the air platform at various values of hovering of the latter.

In chapter ***Principles of Constructing Communication and Control Systems Protected from the Effects of Jamming Attacks for Small-sized Unmanned Aerial Vehicles*** by **M. Kaidenko and S. Kravchuk**, the authors are concerned about the vulnerabilities of small-sized unmanned aerial vehicles (UAVs) in terms of communication and control systems. The research aimed at developing methods for creating high-tech countermeasures is presented. The classification of attacks and hindrances according to various criteria is given. A taxonomy of various types of known attacks that UAVs may be subject to is presented. Intentional interference can lead to a temporary loss of control of the unmanned vehicle, its incapacitation, interception of control of the vehicle, and, as a consequence, its misuse. It is noted that the security of UAVs is a major problem, especially in terms of cyberattacks and radio channel jamming attacks, while there is currently no typical way to effectively counter them. A promising architecture of a communication and control system for a small-sized UAV is presented; its key features and implementation options are considered. The presented architecture is aimed primarily at protecting against jamming and radio channel simulation.

**O. Tsukanov and Ye. Yakornov** ***(Estimation of Motion Parameters of Unmanned Aerial Vehicles of Wireless Sensor Networks Based on the Least Squares Method with a Fractional Taylor Series in a Sliding Window)*** present the scientific principles for improving the accuracy of estimating the motion parameters of quadcopters as elements of flying wireless sensor networks. The authors proposed to use for this the fractional Taylor series and carried out a comparative assessment of fractional polynomials with Chebyshev polynomials and polynomials based on the Taylor series.

## Part III Modern Challenges in Radio Electronics Technologies

**A. Zakharov and M. Ilchenko** ***(Mixed Coupling in Trisection and Quadruplet Bandpass Filters)*** presented a novel approach for the inverse problem for trisection and quadruple bandpass filters (BPFs) with mixed cross-coupling. It allows for a given placement of transmission zeros and known main coupling coefficients to determine mixed cross-coupling $K = Km + Ke$, containing magnetic and electrical components. Based on the obtained solution, it was established that trisection BPF has ten different options for placing two transmission zeros on the complex plane S

$= \sigma + j\Omega$. It is shown that the considered trisection and quadruplet BPFs can have a second-order transmission zero on the $j\Omega$ axis, which provides a deeper attenuation pole at insertion loss curve. With the help of the obtained inverse problem solution, some restrictions are established on the possible options for the placement of three transmission zeros of quadruplet BPF with mixed cross-coupling K14: Transmission zeros cannot be placed on the $\sigma$ axis; two of the three transmission zeros on the $j\Omega$ axis cannot be equidistantly relative to $S = 0$. It is found that to obtain a flat group delay, the transmission zeros must be located in the S plane at the corners of an isosceles triangle, the vertex of which lies on the $j\Omega$ axis, and the sides intersect the $\sigma$ axis. In this case, in addition to the flat group delay, the insertion loss curve has an attenuation pole. Theoretical results are validated with two microstrip quasi-inline trisection BPFs and one stripline quadruplet BPF.

**A. Zakharov and M. Ilchenko** ***(Stripline Combline and Pseudocombline Bandpass Filters)*** discuss the patterns of coupling coefficients between stripline resonators in combline and pseudocombline structures. It was established that there is an electromagnetic coupling between $\lambda/4(\lambda/2)$ resonators in such structures, and they are bandpass filters (BPFs). Between stripline stepped impedance resonators (SIRs), both positive and negative mixed coupling can be realized. This coupling can vary widely by changing the shape of resonators and the gap between them. Moreover, the tuning of coupling is carried out without the use of a conducting pin, as in the case of microstrip resonators. The patterns of changes in the coupling coefficients between stripline SIRs at higher resonant frequencies were studied. These changes have a wave-like (alternating) character. The effect of transitioning the coupling coefficient through zero can be used to expand the rejection band of BPF by suppressing the nearest spurious bandwidth. It was found that the coupling coefficients between stripline resonators, all of whose side surfaces are metallized, depend only on the geometric parameters and are not dependent on dielectric constant $\varepsilon_r$. The dielectric constant only moves the coupling frequencies of an insulated stripline structure, while maintaining the ratio between these frequencies. The measurement data for some combline and pseudocombline stripline BPFs is presented.

**S. Rozenko, S. Litvintsev and L. Pinchuk** ***(Transmission Line Dual-Mode Resonators and Dual-Band Filters)*** presented a novel approach for the problem of optimal synthesis of nonuniform transmission line structural elements with a special arrangement of resonant frequencies, which made it possible to expand the functionality of dual-mode resonators based on them. Optimal synthesis consists in ensuring a given ratio between the resonance frequencies of structural elements with the minimum value of parameter $m = Z_{0\max}/Z_{0\min}$, which is an optimality criterion. The method of parametric synthesis was used for optimization. It leads to stepped-impedance transmission line segments (SILS) with the required arrangement of resonant frequencies and $m = m_{\min}$. Two SILS pairs are synthesized. As result, the new four dual-mode resonators with enhanced functionality are obtained. All dual-mode resonators have an extended stopband between the main dual-mode oscillation and the nearest parasitic oscillation. At $m = 5$, which is the maximum for planar transmission lines, the parasitic bandpass can be located 6.47 times higher than the main bandpass. The dual-mode resonators have reduced dimensions, and their length

shortening relative to a half-wave resonator can reach 0.535. The resonators allow increasing their operating frequencies. The excess factor relative to the resonant frequency of a half-wave resonator can be 2.44. This resonator is very promising for use in dual-band BPF, since it allows us to change the relative position of two passbands in the range $1.46 \leq f_2/f_1 \leq 6.47$. This chapter presents the results of EM simulation of four microstrip filters and the measurement results of two microstrip filters. The results obtained can be used in triple-mode resonators.

In chapter ***Microwave Resonant Structures with Metamaterial Properties as Models of Some Quantum Interference Processes***, **H. Avdieienko, M. Ilchenko, R. Kamaraly and A. Zhivkov** proposed new models based on previously published models of microwave metamaterial structures simulating the Fano resonance. Mathematical expressions have been obtained that describe characteristics such as electromagnetically induced transparency (EIT) and Autler–Townes splitting (ATS). The possibility of using the same models for simulating several resonant processes with a change in a number of their parameters is shown. The analysis was carried out both on the basis of the interaction of even and odd oscillations in microwave metamaterial structures and on the basis of bridge resonant RLC quadrupoles based on lumped elements. Recommendations are given on the use of some special properties of the amplitude and frequency characteristics of microwave metamaterial structures to distinguish between processes of the EIT and ATS types.

In chapter ***Modification of Capon's Method for Several Radio Sources Coordinates Determining by the Shape of the Electromagnetic Wave Phase Front*** by **H. Avdieienko and Ye. Yakornov**, inefficiency of classical Capon's method of bearing angles estimation of radio sources of far-field region with the flat phase front of electromagnetic wave was shown for bearing angles estimation of radio sources with spherical phase front of electromagnetic wave located in the intermediate-field region relatively to the inputs of linear antenna array of radio direction finder. A modification of Capon's method to deal with the spherical electromagnetic wave front is proposed, and some simulation results of bearing angles estimation for several radio sources located simultaneously in the intermediate-field and far-field regions are presented.

The chapter ***Different Approaches for Analytic and Numerical Estimation of Operation Temperature of Cooled Cathode Surface in High Voltage Glow Discharge Electron Guns*** by **I. Melnyk, S. Tuhai, M. Surzhykov, I. Shved, V. Melnyk and D. Kovalchuk** provided analytical and numerical estimation of operation temperature of the surface of cooled cathode in the high-voltage glow discharge electron guns. All theoretical presumptions are based both on analytical and numerical solving of heat partial differential equation. The different constructions of the cathode cooling system are also considered. Obtained analytical solutions are based both on transforming of Boltzmann thermodynamic equation and on considering Bessel and integral exponential functions expansion for nuclear of heat equation. Obtained numerical solutions are based on applying of CAM solidworks simulation software instrument. Experimental estimations of the cathode energetic efficiency for different metals and sort of residual gases are also given. The simulation results are shown, that with power of electron gun till 500 kW and using aluminum as cathode material, for the suitable construction of cooling system, the operation temperature

of the cathode surface is not greater than 200 °C. It is also proving, that difference between the preliminary estimation based on Boltzmann Thermodynamic Equation and numerical calculation using CAM Solidworks is usually smaller than 30%, and simple analytical approach gives the correct value of cathode surface temperature.

We would like to sincerely thank the authors of this collection, because without their hard work of preparing good chapters this volume would not have been successfully prepared.

**And, last but not least, we wish to thank** Series Editor, Prof. Janusz Kacprzyk, Polish Academy of Sciences, Dr. Thomas Ditzinger, Executive Editor, Interdisciplinary and Applied Sciences & Engineering, Ms. Viradasarani Natarajan, Project Coordinator, Books Production, and everyone who has involved in this project from Springer Nature for their dedication and help to implement and finish this large publication project on time maintaining the highest publication standards.

Kyiv, Ukraine
Warsaw, Poland
June 2022

Mykhailo Ilchenko
Leonid Uryvsky
Larysa Globa
Janusz Kacprzyk
Series Editors

# Contents

**Modern Challenges in Information Technologies**

**Towards Role-Based Context-Aware Monitoring Systems** . . . . . 3
Ilja Shmelkin, Daniel Matusek, and Alexander Schill

**Adaptation Consistency of Distributed Role-Oriented Applications Based on the Actor Model of Computation** . . . . . 35
Daniel Matusek, Tim Kluge, Ilja Shmelkin, Thomas Springer, and Alexander Schill

**Ontology-Driven Approach to Scientific and Educational Information Representation** . . . . . 61
Larysa Globa, Rina Novogrudska, Bohdan Zadoienko, and Yu Junfeng

**Approach to Uniform Platform Development for the Ecology Digital Environment of Ukraine** . . . . . 83
Larysa Globa, Stanislav Dovgyi, Oleh Kopiika, and Oleksii Kozlov

**Cloud-Based Technologies for Data Processing in Ukraine: International Context** . . . . . 101
Andrii Shelestov, Bohdan Yailymov, Hanna Yailymova, Svitlana Nosok, and Oleh Piven

**The Comprehencive Approach to Big Data Preprocessing** . . . . . 119
Larysa Globa, Rina Novogrudska, and Mariya Grebinichenko

**Mathematical Models and Informational Technologies of Crop Yield Forecasting in Cloud Environment** . . . . . 143
Leonid Shumilo, Sofia Drozd, Nataliia Kussul, Andrii Shelestov, and Sergiy Sylantyev

**Compulsoriness and Energy Efficiency in the Decentralized World of Smart Things** . . . . . 165
Andriy Luntovskyy, Tim Zobjack, and Mykhailo Klymash

**Standard Model of System Architecture of Enterprise IT Infrastructure** . . . . . 181
Stanislav Dovgyi and Oleh Kopiika

**Comparative Analysis of Object Detection Methods in Computer Vision for Low-Performance Computers Towards Smart Lighting Systems** . . . . . 203
Ivan Matveev, Kirill Karpov, Maksim Iushchenko, Dmitrii Dugaev, Ivan Luzianin, Eduard Siemens, and Ingo Chmielewski

**Optimization of Control Characteristics Using the Information Model** . . . . . 217
Nina Lyubymova, Artem Naumenko, Oleksandr Poliarus, Yevhenii Chepusenko, and Alexandra Lioubimova

**Modern Challenges in Telecommunication Technologies**

**Resilience Improvement by Traffic Engineering Fault-Tolerant Routing in Programmable Networks** . . . . . 235
Oleksandr Lemeshko, Oleksandra Yeremenko, Maryna Yevdokymenko, Amal Mersni, Valentyn Lemeshko, and Mykhailo Persikov

**Research of Automated Control Systems Development Based on "Publish-Subscribe" Technology Over Low-Bandwidth Radio Networks** . . . . . 257
Irina Strelkovskaya and Roman Zolotukhin

**Multipoint Data Transmission Issues in High Bandwidth-Delay Product TCP/IP Networks** . . . . . 277
Nikolai Mareev, Dmitry Kachan, and Eduard Siemens

**Research of the Service Structure Influence on the Sensitivity Indicators of the Queuing System Characteristics with Priorities** . . . . . 297
Leonid Uryvsky and Kateryna Martynova

**Improving the Accuracy of User Location in the Wi-Fi Network Using Complex Spline-Functions** . . . . . 317
Irina Strelkovskaya, Irina Solovskaya, and Juliya Strelkovska

**Principles of Building Modular Control Plane in Software-Defined Network** . . . . . 333
Oleksander Romanov, Mykola Nesterenko, Volodymyr Mankivskyi, and Oleksandr Zhuk

**The Method of Using a Telecommunication Air Platform as a Flying Info-Communication Robots** . . . . . 357
Oleksandr Lysenko, Valery Romaniuk, Anton Romaniuk, Valery Novikov, Valery Yavisya, and Ihor Sushyn

**Wireless Connection of Drones to the Base Station of the Existing Terrestrial Mobile Network** . . . . . 377
Serhii Kravchuk and Irina Kravchuk

**Principles of Constructing Communication and Control Systems Protected from the Effects of Jamming Attacks for Small-Sized Unmanned Aerial Vehicles** . . . . . 399
Mykola Kaidenko and Serhii Kravchuk

**Estimation of Motion Parameters of Unmanned Aerial Vehicles of Wireless Sensor Networks Based on the Least Squares Method with a Fractional Taylor Series in a "Sliding Window"** . . . . . 419
Oleg Tsukanov and Yevgenii Yakornov

**Modern Challenges in Radio Electronics Technologies**

**Mixed Coupling in Trisection and Quadruplet Bandpass Filters** . . . . . 439
Alexander Zakharov and Michael Ilchenko

**Stripline Combline and Pseudocombline Bandpass Filters** . . . . . 469
Alexander Zakharov and Michael Ilchenko

**Transmission Line Dual-Mode Resonators and Dual-Band Filters** . . . . . 499
Sergii Rozenko, Sergii Litvintsev, and Liudmyla Pinchuk

**Microwave Resonant Structures with Metamaterial Properties as Models of Some Quantum Interference Processes** . . . . . 535
Hlib Avdieienko, Mykhailo Ilchenko, Roman Kamaraly, and Alexander Zhivkov

**Modification of Capon's Method for Several Radio Sources Coordinates Determining by the Shape of the Electromagnetic Wave Phase Front** . . . . . 553
Hlib Avdieienko and Yevhenii Yakornov

**Different Approaches for Analytic and Numerical Estimation of Operation Temperature of Cooled Cathode Surface in High Voltage Glow Discharge Electron Guns** . . . . . 575
Igor Melnyk, Serhii Tuhai, Mykola Surzhykov, Iryna Shved, Vitaliy Melnyk, and Dmytro Kovalchuk

**Author Index** . . . . . 597

# Modern Challenges in Information Technologies

# Towards Role-Based Context-Aware Monitoring Systems




**Ilja Shmelkin, Daniel Matusek, and Alexander Schill**

**Abstract** Monitoring IT systems during operation is one of the few methods that help administrators track the health of the monitored system, predict and detect faults, and assist in system repair and fault prevention. However, existing monitoring solutions still suffer from several trade-offs, as current implementations impose architectural constraints on monitored systems that result in less flexibility in deployment and operation. While there are excellent monitoring systems available for some application areas, others are not adequately supported. In addition, most monitoring software is specialized to work with specific data formats, protocols, and data collection mechanisms, further limiting its flexibility. The role-based approach to modelling and implementing software promises an intuitive way to increase flexibility in modelling and implementing information technology systems. Coupled with control loop technology from the field of self-adaptive systems, we created a reusable framework of static role-playing building blocks to overcome these limitations. In this chapter, we present and discuss an ongoing research project on new concepts for creating flexible role-based monitoring systems that function in most application domains. For that, we discuss the results of a study that compared the flexibility of a variety of monitoring systems and compare the flexibility of our concept to them based on 11 criteria. We illustrate our concept by presenting and discussing a role-based model for control loop components. We further discuss our implementation of the concept as well as how it can be applied to different problem domains. Finally, we discuss the results of a quantitative evaluation of the role-based monitoring approach and present an outlook of future tasks that need to be solved in this research project.

I. Shmelkin (✉) · D. Matusek (✉) · A. Schill (✉)
Technische Universität Dresden, 01062 Dresden, Germany
e-mail: ilja.shmelkin@tu-dresden.de

D. Matusek
e-mail: daniel.matusek@tu-dresden.de

A. Schill
e-mail: alexander.schill@tu-dresden.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_1

**Keywords** Self-adaptive systems · MAPE-K · Roles · Control loop · Feedback loop · Modeling

## 1 Introduction

The use of information technology (IT) systems continues to expand into new areas of application. With each new area, new challenges arise that researchers and developers must overcome. During operation, IT systems tend to malfunction or behave unexpectedly, requiring an administrator to intervene to detect and correct errors. Unfortunately, modern systems are often too complex and produce too much data for humans to understand why unexpected behavior and malfunctions occur. For many years, monitoring systems have been used to help humans identify faults, detect root causes, troubleshoot, and prevent failures [2, 11, 13, 16, 19–22, 29, 32].

In a recent study [25], we found that monitoring systems generally work very well for certain application areas, such as server monitoring, operating system monitoring, and application monitoring. However, we also found that most monitoring systems only support a narrow range of functions and protocols that are useful for the particular application area. As a result, there are many specialized monitoring systems available on the market that allow monitoring of specific applications, operating systems, or servers through predefined workflows and fixed protocols. For individuals and companies that have a larger infrastructure of IT systems that need to be monitored, it is currently very difficult to find a monitoring system that supports all of their requirements. This is even more true for individuals and companies that use systems from past decades that are not compatible with the features of current monitoring systems. Furthermore, when proprietary protocols and highly specialized workflows must be used for monitoring, modern monitoring systems are not flexible enough to be adapted to function properly in such cases. This results in the need to use multiple monitoring systems or, in certain use cases, having no suitable monitoring system at all available. The need for a flexible monitoring system that can be deployed in any application domain and whose functions are fully customizable, allowing the user to decide on the used protocols, data collection strategies, implementation of analysis functions, alerting strategies and notification endpoints, exists more than ever.

This chapter provides an overview of our current research on a role-based context-aware monitoring system with a control loop structure that is fully flexible with respect to all of the above criteria [24–26]. To this end, we begin by presenting an overview of various techniques for designing flexible systems in Sect. 2. Also, we highlight how flexibility can be achieved with the right implementation of a monitoring system's functions. Afterwards, we discuss our comparison of fifteen monitoring systems from different decades to gain insights into the structure of monitoring systems and how they provide functionality to the user. We will then discuss why we identified roles in context as an appropriate technology for designing and developing a fully flexible monitoring system. Paired with roles, we explain why

a feedback loop design is beneficial for monitoring systems and discuss how we combine such a design with the role concept in Sect. 3. We discuss how a role-based control loop can be used to create a flexible monitoring system and discuss how this concept can be implemented and applied to different application domains. Along this, we discuss a scenario and system structure from one such application domain for the evaluation of this concept in Sect. 4. Finally, we discuss the next steps to be solved in our research project and possible future work in Sect. 5.

## 2 Related Work

We have found that monitoring systems are typically inflexible because there is strong cohesion between the system and its application domain [25]. This is largely due to the fact that developers specify, for example, fixed workflows, data collection strategies, and analysis algorithms as well as limited alerting and notification capabilities. With that, most monitoring systems are tailored to one specific application domain that fits the provided set of functions but cannot be adapted towards working with other application domains. In this section we will discuss two approaches on creating more flexible (monitoring) systems. First, we will discuss techniques from software engineering, i.e., software design patterns, that boost flexibility of the software by allowing to add new or changed behavior during its life cycle. Second, we discuss how flexibility of (monitoring) software changes, depending on how its features are implemented and provided to the user.

### *2.1 Software Design Patterns*

In software technology, *design patterns* provide a method to overcome different reoccurring software design problems. Given a particular application, different design patterns may be applied to the structure of the software to boost certain aspects like: re-usability of code, dynamic behavior, software flexibility, decomposition of complex algorithms, and many more. While those patterns usually focus on providing benefits for certain aspects of the software, they also inflict detrimental on others (e.g., comprehensibility, software size, etc.). Thus, it is necessary to decide on a case-to-case basis, which design patterns should be used for specific software and if some may be used in conjunction to provide even more benefits.

First of all, design patterns are roughly divided into *inheritance-based* patterns and *delegation-based* patterns. While inheritance models an "*is-a*" relationship, delegation models the *behavior of an entity.*

**Inheritance-Based Patterns**

Inheritance allows a *child class* (also: derived class) to inherit the attributes and methods of its *parent class* (also: base class or super class). A child class can also define new *attributes* and *methods* as well as *overload* (i.e., overwrite) the inherited properties. Using inheritance is beneficial when a child class is a natural extension of its parent class (e.g., a computer is-an electronic device). Applied to monitoring systems, we can create a generic monitoring class that defines all common attributes and functions of a monitoring system. The case-specific implementation is then created by an inheriting child class (e.g., a *mailserver monitoring system* is-an *application monitoring system*, which is-a *monitoring system*). Naturally, this inheritance chain can get complex very quickly. Especially when small nuances change between children on the same level of inheritance, *combinatorial explosion of inheriting classes* occurs.

*Strategy Pattern*

The *strategy pattern* [9] (also: policy pattern) enables the *decoupling of the client and the algorithm used by the client* [8]. This is achieved by encapsulating different strategies for aspects of a class in *behavior interfaces.* In this way, a monitoring system can have multiple behavioral interfaces for all methods, which can change their *strategy* (i.e., their behavior) from case to case. Instead of creating a new child class for a new use case, we can use the strategy pattern to choose between different strategies that fit the use case. When a new strategy is needed, it is created as a child class of the respective behavior interface. This leads to massively improved code reusability and maintainability. This is even more true the higher the level of specialization of each behavior interface, which means that creating one interface for the entire monitoring process provides less reusability and maintainability than creating many interfaces for many smaller parts of the monitoring process. This phenomenon enables the intuitive application of the *template method pattern.*

*Template Method Pattern*

The template method pattern [8, 9] allows the methods of an abstract class to be divided into several smaller *helper methods*, each of which performs a particular step of the overall algorithm. For example, the monitoring process as such can be divided into several equally important tasks such as *data collection*, *data analysis*, *alert and notification management,* and *system adaptation.* Combined with the strategy pattern, we can create a behavioral interface for each of the helper methods. In this way, strategies are now provided for a fraction of the entire algorithm.

The more helper methods we use, the smaller each of them becomes. A large number of helper methods allows the creation of many, very specialized strategies. This increases the flexibility of a (monitoring) system, as support for new use cases can be added by using existing strategies or implementing new ones if no suitable strategy exists.

Developers must balance the number of helper methods (i.e., the *granularity*) because the users of the system must be aware of all of them. A system that is too fine-grained may overwhelm users, while one that is too coarse-grained will not provide the same level of flexibility and code reusability, as large portions of the system will need to be re-implemented to support a new use case. However, with inheritance-based patterns, a large number of fine-grained strategies on a multi-level inheritance chain almost certainly result in incomprehensible models and large software.

Monitoring systems have the peculiarity that small changes in the monitored system (e.g., supported *transport protocol*, changed *scraping intervals*, change in *data format*, etc.) already imply that the current configuration of the monitoring system is no longer functional. Over the long term, this implies that a flexible inheritance-based monitoring system has to maintain many different strategies, thus, it has to be large and incomprehensible.

**Delegation-Based Patterns**

Software technology also provides another option to achieve code reuse and also flexibility in software that is as powerful as inheritance. With *delegation*, we can define *delegate classes*, that have a reference to their *original class* by using composition [14]. During run-time, given specific circumstances (i.e., a *context*), the delegate classes perform actions that fit the purpose of the original class in the given context. Thus, with delegation, two objects are involved in handling a request: *a receiving object delegates operations to its delegate* [9].

With the concept of delegation, again, we can define an abstract interface for the components of a (monitoring) system. An implementation of this interface, the original class, holds a reference to a delegate, which also implements the same interface. Different delegates of the same abstract interface model different behavior of the interface, e.g., one delegate handles *data acquisition* via Google RPC (gRPC) while another may use the Simple Network Management Protocol (SNMP) or the Hyper Text Transfer Protocol (HTTP).

This relationship between the original class and the delegates can be understood as a role-playing relationship, in which the delegate is played by the original class in a given context. Since delegates encapsulate *context-specific behavior*, we achieve the same flexibility compared to a combination of the inheritance-based strategy and template method patterns. However, the size of the respective software models is smaller with delegation, thus, comprehensibility is increased in a monitoring system scenario, as the behavior of a class is bound to a given context. In the special case of monitoring systems, using delegation is favorable compared to inheritance

because the use of monitoring systems often occurs in multiple contexts (i.e., different monitored systems).

**Role-Based Modeling**

To emphasize the role-playing relationship that occurs between an original class and its delegates (and to make the resulting models more intuitive and natural), the *role-object pattern* was introduced [4], a combination of the *decorator pattern* [9] and the *product trader pattern* [3].

*Role-Object Pattern*

The role-object pattern makes the *role-playing relationship* explicit within an *object-oriented* model using delegation. With that, abstract interfaces do not define any implementation state but only component-unspecific communication operations as well as a minimal protocol for managing roles. Roles are modeled as child classes of a common abstract parent class. While the common abstract parent class is never instantiated, instead, a role is chosen that fits the provided context and instantiated during run time. Each role can define additional operations and attributes as needed in a specific context, e.g., the monitoring of an application via different protocols, similar to what we achieved with pure delegation in the previous subsection. With this pattern, we achieve a context-specific view on the components of our monitoring system and are able to switch out roles at run time that are associated with our core object if the context of our system changes.

Summarized, the role-based approach is superior to other modeling approaches for increasing flexibility in *monitoring systems* because:

1. It allows handling different context-specific versions of a component (a class) without combining all context-specific operations and attributes in one common interface. This results in the component being well-focused on the essential state and behavior of the modelled key abstraction [4].
2. It allows handling available context-specific versions dynamically during run time by adding or removing roles.
3. New context-specific versions, i.e., roles, can be added easily while preserving the component's implementation. This allows supporting a variety of different monitoring use-cases without interference between the use-cases (i.e., separation of concerns).
4. Compared to inheritance-based patterns, it has greatly improved comprehensibility, as models are less complex through avoidance of combinatorial explosion of classes. Context-specific operations and attributes are encapsulated within multiple roles instead of one single class or class hierarchy.
5. Compared to pure delegation, it makes use of roles as an intuitive concept for modelling *context-specific* operations and attributes.

Unfortunately, the role-object pattern offers no method to make context explicit within models. This is a downside that all object-oriented models have in common. As monitoring software acts in numerous different contexts, i.e., one per monitoring use-case, making context explicit offers large benefits for comprehensibility. Pure

role-based technologies, as introduced below, allow to make context explicit, hence, they are superior compared to an object-oriented representation of roles.

*The Notion of Roles in Context*
The concept of roles in software modeling and development was originally introduced by Bachmann and Daya [1] and later refined by several researchers [10, 27, 28]. Its concept describes that objects can be separated in their static attributes, which represent their very *nature* and the *roles* they play within a specific *context*. The notion of roles further describes that static *naturals* play context-specific *roles* in a given *context*. When the context of an object changes, its context-specific purpose (the role it plays) also changes, while its static nature always remains the same. Pure role-based technology works similarly to the delegation-based role-object pattern.

We have found that this concept allows to remove the strong cohesion between a monitoring system and its monitored system in a given use case. It allows decoupling the abstract architecture of monitoring systems from their domain-specific purpose to provide separation of concerns and full flexibility. This means that for each individual use case we can create a set of roles that implement exactly the functionality needed to support the use case. As a result of this consideration, a concept and architecture for a monitoring system that is based on roles is described in detail in Sect. 3.

## 2.2 Flexible Feature Implementation

Apart from flexibility resulting from the design of software systems, we can also classify flexibility at the *level of provided functions*. That is, some monitoring systems provide functions in such a specific way that their application is limited to a narrow range of systems, while others provide the same functions in a *broader, more general, and more flexible way*. In this section, we discuss the results of a recent study [25], in which we compared the flexibility of 15 different monitoring systems based on 11 flexibility criteria. Consequently, we will combine the role concept with our findings on the implementation of flexible functions and flexible software design.

**Systems Flexibility Comparison**
Monitoring systems today accomplish much more than just monitoring a set of configured values over time. They also provide data analysis functions and allow action to be taken based on the analysis results.

Due to the enormous number of different monitoring systems on the market, there are many ways in which certain functions are made available to the user. For example, while one monitoring solution uses gRPC to transmit monitoring values, others may use SNMP, HTTP, as highlighted earlier. In addition, some systems rely on the monitored system to transmit values at regular intervals, while others retrieve values from the monitored system independently. Analysis capabilities can also be implemented in a variety of ways, with simple threshold analysis being supported by most monitoring systems, while other monitoring systems sporadically offer their

**Table 1** Evaluation of flexibility categories of IT monitoring systems. ○ = not flexible; ◐ = partially flexible; ● = flexible, □ = not supported; ■ = supported

| Category | Flexibility Criterion | [9] | [12] | [17] | [10] | [26] | [19] | [18] | [16] | [2] | [5] | Nagios | Prometheus | Elastic Stack | Influx Stack | Sensu | Role-based Monitoring Concept |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | 1. Data Formats | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ◐ | ○ | ● | ● | ● | ● |
| | 2. Data Collection | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ● | ◐ | ◐ | ● | ◐ | ● | ◐ | ◐ | ● |
| A | 3. Data Analysis | ○ | ○ | ○ | ○ | ◐ | ◐ | ○ | ○ | ○ | ○ | ○ | ○ | ◐ | ◐ | ◐ | ● |
| P | 4. Alerting | □ | □ | ■ | □ | ■ | ■ | □ | ■ | □ | □ | ■ | ■ | ■ | ■ | ■ | ■ |
| | 5. Notification Sinks | ○ | ○ | ◐ | ○ | ◐ | ◐ | ○ | ◐ | ○ | ○ | ◐ | ◐ | ◐ | ◐ | ◐ | ● |
| | 6. Autonomic Behavior | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| E | 7. Execution | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ◐ | ○ | ● |
| K | 8. Database System | ○ | ◐ | ◐ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● | ● |
| | 9. Knowledge Interface | ○ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ● | ◐ | ● |
| Other | 10. External Tooling | □ | □ | □ | □ | □ | □ | □ | □ | □ | □ | ■ | ■ | ■ | ■ | ■ | ■ |
| | 11. Plugin Extension | □ | □ | □ | □ | □ | □ | □ | □ | □ | □ | ■ | □ | ■ | ■ | ■ | □ |

users advanced capabilities such as reasoning, machine learning, outlier and anomaly detection, and more.

There is no compact way to compare all features with all their different expressions. Therefore, we have summarized the different features and their expressions as different criteria. Each criterion expresses an important functional complex of monitoring systems. As an example, we can consider the criterion "data formats". It represents several possible expressions (e.g., metrics, events, log files, etc.). In addition, each expression usually has a number of different implementations, most of which are used only by one particular monitoring system.[1]

We further selected 15 different monitoring systems to analyze and evaluate with respect to these criteria. The systems were selected primarily on the basis of whether they represented a technical innovation, which is why most of them were described in patents. We also tried to select systems from different time periods, since one goal of the classification was to see how monitoring systems have changed over time. In particular, we wanted to see if there has been a shift over time toward more flexible functional implementations.

We continue with a brief description of each criterion. A complete description of the criteria can be found in our recent study [25]. Furthermore, Table 1 depicts the classification of the 15 monitoring systems that we evaluated during our study.

*Data Formats*

Different data formats can be used to collect data from the monitored system. While specific systems supporting only one use case [2, 6, 13, 21, 22] and systems from past decades [11, 16, 20] mainly use proprietary formats, modern systems (e.g., Nagios, Prometheus, Elastic Stack, Influx Stack, and Sensu) use more granular ways of formatting data, e.g., metrics, log files, events, and checks. Today, there are some efforts to provide well-defined format definitions (e.g., OpenMetrics), though the monitoring community has yet to fully commit to standards for them.

[1] Note that we did not include all possible expressions in our evaluation.

*Data Collection*
In addition to the different formats, there are also different ways in which the data can be collected. While early systems and highly specific systems used only hardware agents to collect data from a system [11, 13, 16, 21, 22], with advancements over time, systems more and more used *software agents*, which can be deployed on various operating systems [2, 19, 20, 32]. The modern state-of-the-art systems (e.g., Nagios, Prometheus, Elastic Stack, Influx Stack, and Sensu) usually support both ways, software and hardware agents. However, support for hardware agents is usually achieved through the SNMP protocol. In addition, systems differ in their strategy for triggering data transport, i.e., some systems pull data from the agents, while in other systems the agents periodically forward the data to the monitoring system. There are also current efforts to standardize data collection (e.g., OpenTelemetry), but the monitoring community has not yet fully committed to these contributions.

*Data Analysis*
This criterion summarizes commonly found analysis functions that are applied to monitored data. The systems on our list are supporting a subset of simple threshold analysis, outlier and anomaly detection, machine learning, and more. Highly specialized systems [11, 13, 16, 21, 22] use *fixed hardcoded algorithms* for data analysis. More recent and more general systems used *primitive rule-based approaches* [19, 32], while state-of-the-art systems usually rely on their own *expression language* to define analysis tasks (Nagios, Elastic Stack, Influx Stack, and Sensu).

*Alerting*
An important feature of monitoring systems is the ability to notify an entity when the Analysis component concludes that the state of the monitoring system is out of order. Again, modern systems use an expression language to define *alerts*. Typically, this is the same language as for the definition of analysis tasks in state-of-the-art systems (Nagios, Elastic Stack, Influx Stack, and Sensu). Alerts are often tightly coupled with analysis features, i.e., alerts are defined over the outcome of analysis tasks. Older and highly specialized monitoring systems use fixed hardcoded alerts instead of user-defined ones.

*Notification*
Traditionally, systems used local logging and E-Mails to send notifications based on triggered alerts. More recently, additionally, a variety of instant messaging providers (i.e., Slack, Pushbullet, etc.) can be used as well as different database drivers to easily notify any device or human.

*Autonomic Behavior*
In more recent systems (e.g., Nagios, Prometheus, the Elastic Stack, Influx, and Sensu), notifications cannot be sent exclusively to be read by humans, but also to be processed by software. For that, systems allow sending notifications to an API via HTTP/S, allowing the API to perform autonomic actions based on the notification. However, there are also different types of APIs that are commonly available in hardware like switches and routers, e.g., SNMP's extend functionality.

*Execution*

While, most of the time, external APIs are used to trigger actions based on received alerts, there is also the option to use an execution environment that is directly included within the monitoring system. For that, either the monitoring system can locally execute actions or remotely trigger actions by using software agents.

*Database Connectivity*
Naturally, every monitoring system produces data that should be persisted for a period of time. This data consists mostly of monitored data from the managed system but also includes system configuration, user accounts, alerting rules, notification configurations and more from the monitoring system. Due to the nature of gathering data periodically essentially always forming a time-series, databases that support time-series are an intuitive choice. Besides that, both the monitoring system and the monitored system usually also produce relational- or key-value data. This creates the necessity to deal with different data types during operation, thus, ideally, a monitoring system should support different types of databases in cooperation.

*Knowledge Interface*
Gathered data, which is persistently saved within a monitoring systems database, may be used for a variety of tasks associated with monitoring, analysis, planning, and execution (these stages are described later in more detail). Intuitively, an interface for each stage to query data from the database of the system should be provided.

*External Tooling*
Such dynamic access to data is especially useful for adding the support to external tooling. In modern monitoring ecosystems, the application of data science techniques or graphing, for example, is often achieved through third-party tools. It is therefore very useful, to provide the same options for querying data to externals.

*Plugin Extension*
Some monitoring systems (e.g., Nagios, Elastic Stack, Influx Stack, and Sensu) support the extension of their function set via *plugins*. The developers of Nagios were the first to recognize, that a system should provide flexibility by design. Therefore, a powerful *plugin engine* was included within Nagios, that was used by countless developers to create checks for close to anything in the past 20 years. While a plugin engine usually provides a variety of possible additions, developers are still limited by restrictions from the parent system, e.g., data formats, data collection mechanisms, and protocols. Nevertheless, we identified that plugins are the most powerful tool to make a monitoring system support more features.

*Systems Classification*
Evaluating our classification as a whole, we can see that the flexibility of modern monitoring systems is far greater than the flexibility of past systems, except for Nagios, which was introduced already in 1999, yet provides above-average flexibility due to its heavily integrated plugin system.

We can also see that, regardless of modernness, systems typically have good flexibility regarding their supported interfaces to the system's knowledge and their alerting capabilities. Flexibility in recent systems is also decent in the interfaces that they provide for external tooling and the support of a plugin engine.

We see small deficits in flexibility regarding the supported data formats, data collection methods, and autonomic behavior.

Finally, we see big deficits in flexibility regarding the data analysis capabilities, the capabilities to execute actions on local and remote systems, as well as the support for different database technologies.

We identified that the flexibility of modern systems is provided mostly through the use of plugins but can also be achieved through hardcoding functions that are specifically made to support multiple use-cases. However, neither of both does allow for each possible concern to be extensible.

**Roles for Flexible Feature Implementation**

Roles behave similarly to plugins from a function-centric perspective. A user can provide a set of roles that cover the functional requirements of a role-based monitoring system for a desired use case (i.e., context). While currently used monitoring systems use plugin engines that are limited by functional constraints of their parent systems, roles do not have this limitation. Instead, roles provide fully flexible functionality within their architectural framework. There is no limitation in what a role can accomplish as long as it does not violate the semantic architecture of its system. Reasoning from that, a role-based monitoring system should have dedicated components for its natural tasks, such as *data collection* and *filtering*, data analysis, *alert* and *notification management*, and further. Also, this means that roles that are played by components at runtime should only be a natural extension of the components' tasks, e.g., the monitoring component should only take care of data collection and filtering. This requires a well-defined architecture of role-playing components and communication interfaces, with each component having a well-defined and delimited set of tasks.

In the following section, we will combine our insights on flexible software development and flexible feature implementation with such a well-defined architecture of components to provide a basis for implementing a flexible role-based monitoring system.

## 3 The Role-Based Monitoring Approach

Our results show that a role-based system design combines the flexibility of delegation-based modeling with explicit contexts in which roles provide context-specific functionality that perfectly fits a given use case. In addition, roles behave similarly to plugins, but do not have to adhere to the constraints of their parent system because roles have no restrictions on the functions they can provide. However, we have also learned that roles rely on a well-defined system architecture with dedicated components that set limits on the range of functions each role can provide.

In our study [25] of monitoring system flexibility, we found that modern monitoring systems often use a common system architecture in which they have dedicated

architectural components for tasks such as data acquisition, data processing and analysis, alarm and notification management, and data persistence. We have found that this common architecture shares many similarities with a control loop structure that is commonly used in the field of self-adaptive systems [30]. The most prominent representative of such a self-adaptive control loop is the MAPE-K control loop. In this section, we will discuss how role-based technology can be used together with the MAPE-K control loop. To this end, we first introduce the basics of the MAPE-K control loop. We then discuss how control loop components can be modeled as role-based components using the Compartment Role Object Model. Consequently, we discuss how this concept can be implemented using the role-based programming language ObjectTeams Java. Finally, we discuss three application domains and how this concept can be applied to support their monitoring and management.

## *3.1 The MAPE-K Control Loop*

Since *self-adaptive systems* were recognized as a way to tame the growing complexity in software systems, *control loops* have been a popular approach to engineer them [7, 15]. The *MAPE-K control loop* is the most prominent example of a control loop. Its concept was originally outlined by Kephart and Chess [15] (as the *autonomic manager*) and later marked out by IBM [12].

The control loop embodies the managing part, or phrased differently, the *managing subsystem*, whereby the *managed subsystem* embodies the actual domain-specific function. Together, they form a self-adaptive system [31]. In the context of this chapter, a monitoring system (not to be confused with the *monitoring component*) can act as the managing subsystem, while the managed subsystem consequently is the monitored system.

The MAPE-K control loop consists of five components whose functions are presented below. The description is supported by Fig. 1, which shows the five components (i.e., the *managing subsystem*), a monitored system (i.e., the *managed subsystem*) and their environment.

It should be noted that the communication order within the control loop is fixed. This means that the monitoring component always forwards its produced data to an analysis component. Accordingly, the analysis component acts as a recipient of the monitoring component's data. Similarly, the planning component acts as a consumer of the analysis component's data, the execution component acts as a consumer of the planning component's data, and the monitored system receives adjustments from the execution component. Finally, all components can exchange data with the knowledge component. The communication order cannot be reversed or altered, but an addition can be introduced within each stage of the loop where components of the same type can communicate with each other (see, for example, Fig. 1). This is particularly useful in distributed scenarios and will be used later in this section.

A brief introduction to the components of the control loop follows. In the meantime, we will map the functions that are usually provided by monitoring systems to

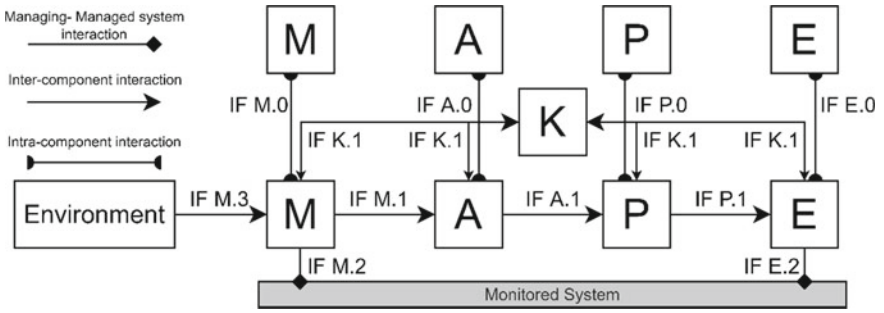



**Fig. 1** Generic MAPE-K control loop [12] with communication interfaces [26]

the components of the control loop to outline our concept, which will be presented later in this section.

**Monitoring Component**

The *monitoring (M) component* aggregates data from the underlying managed subsystem, the managing subsystem (via saved knowledge), other monitoring components, or the surrounding environment. The acquisition may be supported by two mechanisms. Firstly, the managed subsystem may periodically send data to the monitoring component (the *push-strategy*). On the other hand, the *pull-strategy* allows the monitoring component to query the managed subsystem for data. The pull-strategy is usually assisted by probes (also called sensors) on the managed subsystem. Probes are lightweight programs, that open an interface to the managed subsystems internals, so that the monitoring component may query data about those. In the domain of monitoring systems, such probes are usually called agents.[2]

Every monitoring software has a dedicated component to gather data from the monitored system or the systems environment, that other than "monitoring component" is usually called either acquisition- or retrieval component. The data collection can be realized in different ways, i.e., via a hardware agent that is installed in the monitored system (e.g., [11, 16]), software agents (e.g., [19, 20, 32]), Elastic Stack,[3] Prometheus,[4] etc.) that are deployed to acquire a systems state, or specific sensors that use proprietary communication to gather sensor data (e.g., [13]). The categories subdivide again with different protocols (e.g., the Simple Network Management Protocol, message queues, etc.) and methods for data representation (e.g., events, metrics, log files) being available.

**Analysis Component**

The algorithmic processing of monitored data is done by the *analysis (A) component*. The goal is to gain knowledge about the state of the managed subsystem and decide,

[2] The terms agent, sensor, and probe can be used interchangeably.

[3] https://elastic.co - Elastic Website.

[4] https://prometheus.io – Prometheus Website.

whether the system state is valid according to high-level system goals. Such goals may be as obvious as limiting the amount of consumed resources or fulfilling performance constraints. However, there are countless approaches of how to model system goals (e.g., thresholds, temporal logic, probabilistic quantification), most of having their own, narrow application area with induced benefits and shortcomings. In any of those cases, the analysis component produces the so-called adaptation decision as input for the planning component.

Within the analysis component of a monitoring system, the data that is gathered by the monitoring component can be analyzed with different algorithms. Most monitoring software nowadays supports at least simple threshold analysis, with advanced analysis features like machine learning, control theory, outlier detection, and reasoning being also included in contemporary systems.

**Planning Component**
With the adaptation decision from the analysis component, the *planning (P)* component will branch in one of two resolutions.

- If the adaptation decision is *negative* (i.e., the managed subsystem's state is valid according to the high-level system goals), the planning component remains idle.
- If the adaptation decision is *positive*, an *adaptation plan* will be calculated, and forwarded as input of the execution component.

The adaptation plan contains a sequence of actions, that, when properly executed, transition the state of the managed subsystem to again fit the high-level system goals. The techniques used to calculate an adaptation plan depend heavily on the application area of the managed subsystem.

Most contemporary monitoring software natively includes an alerting engine that allows to send out notifications to various sinks (e.g., E-Mail, instant messaging, APIs, etc.). The naive usage of such an engine allows the user of the system to define rather simple actions that take place when a certain condition is met. By sending notifications to an API, however, there are no limits on what a notification can cause, e.g., it could kick off a chain of actions on a remote system to correct anomalous behavior, similar to what a self-adaptive system is supposed to achieve [30].

**Execution Component**
The *execution (E) component* of a MAPE-K control loop is responsible for performing every action that is listed in the adaptation plan that it receives as input. Compared to the components for monitoring, analysis, and planning, the conditions under which the execution component operates differ. Where the former ones may abstract away substantial parameters and specialties of the managed subsystem while performing their task, the latter is fundamentally bound to them. In extreme cases, monitoring, analysis, and planning can be agnostic of the internals of the managed subsystem, while the execution has to pay attention to consistency, timings, communication delays, errors, and fallback-strategies.

While most monitoring systems include the monitoring, analysis, and planning components, the execution of planned actions today is commonly outsourced to

external APIs (except for the InfluxStack[5]), that receive notifications from planning to then execute actions on the local or remote machine. The more incidents a complex system can solve on its own, the better it is, as the maintenance effort is reduced. That such features are rare to find is a shortcoming of contemporary monitoring systems. With a concept based on the MAPE-K control loop components, this shortcoming can be lifted.

**Knowledge Component**

Ideally, in every iteration of the MAPE-K control loop, new information about the managed subsystem is gained. The majority of this information is referred to as knowledge and is managed by a dedicated *knowledge (K) component*. By the original definition, the knowledge component is an implementation of a registry, dictionary, database, or another repository that provides access to knowledge according to the interfaces prescribed by the control loops architecture [12]. Due to the fundamental characteristics of adding new data to a database on every iteration of a reoccurring loop, this definition may be naively narrowed down to a database that allows the handling of time-series data, however, this does not always apply. In cases, where the managed system produces, e.g., relational data, it is beneficial to keep the ability to save such data in the knowledge component of the managing subsystem, hence, multiple databases in conjunction should be used in such cases. The purpose of the knowledge component in the managing subsystem is to aggregate data from the other four managing components and allow querying of this information anytime to improve the managing performance or assist the managing processes in general.

Every monitoring software comes either with an included database (e.g., ELK stack, Prometheus, Influx Stack, Nagios RRD, etc.) or with an interface to contemporary databases such as InfluxDB for time-series data, for example. Most systems use their individual database not only for monitored data but also for saving defined analysis tasks, alerts, user accounts, and configuration. Hence, such a database in monitoring systems maps intuitively to the knowledge component of the control loop.

**Communication Interfaces in Control Loops**

We identified that the components of the control loop, regardless of their actual deployment and the grade of distribution, always use fixed communication paths. In an earlier publication [26], we introduced a set of interfaces that describe those communication paths. That are, the communications between components of the same type (suffix 0.0, see Fig. 1), between components in order of the sequence of the control loop (suffix 0.1), between the control loop and the monitored system (suffix 0.2), and between the control loop and the system's environment (suffix 0.3). Figure 1 shows the concrete naming and placements of the interfaces within a MAPE-K architecture. We will use those interfaces as an important part of our concepts for communication primitives.

[5] https://www.Influxdata.com/ - Influxdata website.

### 3.2 Modeling of a Role-Based Control Loop

The MAPE-K control loop has been the subject of research for some time, as shown by several studies [17, 23] and individual contributions [26, 31] which have added their own extensions to its basic definition. In this context, we would like to point out that the concept of the control loop can be understood differently depending on the field of application.

First, when using run-time adaptation of a managed element, which can be an IT system on its own, the control loop can be understood as an external managing part of a specific system. While the managing part (i.e., the control loop) and the managed part (i.e., the domain-specific purpose of a system) are separated, they still work towards a common goal, and most of the time, they were specifically developed to do so. In such cases, we talk about self-adaptive systems, as described in detail by Weyns [30].

A different understanding of a control loop as the managing layer becomes obvious when each, the managing and the managed part are individual software systems that can function without each other but can be used in conjunction to allow *autonomic monitoring and management* of the managed system. Here, the architecture of the control loop is used to create a *reusable* and *maintainable* structure for a managing layer that is not only working with one specific managed system but functions with a variety of those. While not stated explicitly, this architectural style for IT systems can be found for a long time in monitoring systems [25], as explained before.

Using the insights from Sect. 2 on related work, we found that the notion of roles can be used to intuitively increase the flexibility of a control-loop-based monitoring system by allowing control loop components to play roles in a given context. In this way, support for many different monitored systems can be achieved. In the following, we will discuss this concept of role-playing control loop components.

**Concept and System Structure**

For a system that has a MAPE-K control loop architecture, first, we can specify all control loop components and their interfaces as the static parts of the system. As discussed earlier, static parts of a system can be described as *naturals*. The natural components are each specified with a set of fields that hold the interface objects as needed by each respective component type, as depicted in Fig. 1. Furthermore, each of the components holds a *callback()* method that triggers a *three-phase protocol* when activated. The *callback()* method can be activated differently, depending on the use-case (e.g., in a monitoring natural, it could be potentially activated by an external agent over IF M.2 (see Fig. 1), in the case when the agent pushes data, or by a timer within the component. The specifics of each phase are not determined by the natural but by the *roles* that are played at run time. Below, we explain the three phases. Please see Fig. 2 for additional reference.

1. The *retrieval phase*. When triggered, the component starts collecting data from interfaces that allow for incoming data (i.e., M.0, M.2, M.3, K.1). All received data are temporarily saved within the respective interface objects.

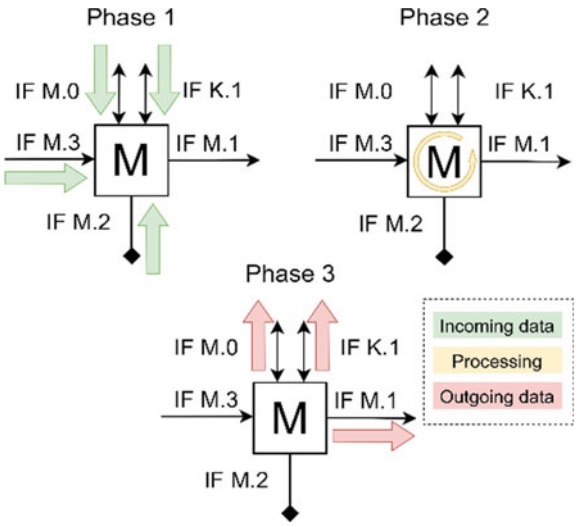



**Fig. 2** Visual representation of the retrieval-, the processing-, and the finalization phase of a monitoring component from the role-based monitoring solution

2. The *processing phase*. When triggered, all temporarily saved data within the interface objects are pulled to the component and processed, e.g., the monitoring component takes care of organizing and filtering monitored values as well as labeling them for further transmission.
3. The *finalization phase*. When triggered, all processed data that are labeled for transmission are sent to the next component in sequence over the respective interface (i.e., M.0, M.1, K.1). After successful transmission, all non-persisted data are deleted from the components' internals so that the component is ready for the next iteration of the loop.

Furthermore, all *interfaces* that are part of a component are also modeled as *naturals*. Each interface consists of three fields (sender information, receiver information, and a dataBuffer for saving temporary data) as well as one or two methods, depending on the interface, one that handles receiving data and/or one that handles sending data. Combined, the model describes a total of 16 different naturals, five for the components of the loop and eleven for all possible interfaces.

The naturals remain agnostic about the specifics of their concrete use case (i.e., their context). All context-specific details are added with roles that are played by the components and the interfaces during run time.

Consider Fig. 3 for a representation of the naturals that are associated with the monitoring functionality, how they are connected, and how a subset of them plays a role in the context *ServerMonitoring*. Between different use cases, naturals always stay the same, however, a new context and its contained roles have to be created. In this case, the monitoring natural plays the role *ServerMonitor* in which its three phases are overloaded with new behavior. Furthermore, the system needs the interfaces M.2

for retrieving performance data, a database connector (K.1), and a connection to its analysis component (A.1), for which three more roles exist in the context.

In Fig. 3, we use the role-based modeling notation Compartment Role Object Model (CROM) [18] as it allows modeling naturals and the roles that they play within a specific context. In CROM, context is modeled as a compartment object, while naturals are modeled as natural types and roles as role types.

With this role-based control loop for monitoring systems, the user has full control of

1. the *data formats* that are used for interactions between the monitored system and the monitoring system as well as between the components,
2. the *data collection mechanisms*, i.e., the use of hardware agents or software agents and if the push- or pull strategy should be used,
3. the *data analysis algorithms* that should be used to decide about the state of the monitored application,
4. how *alerting* is implemented, as there are no restrictions in which *notification providers* may be implemented,
5. how the *execution of actions* is implemented, enabling *autonomic behavior* in local and distributed environments, e.g., via RPCs or existent APIs.

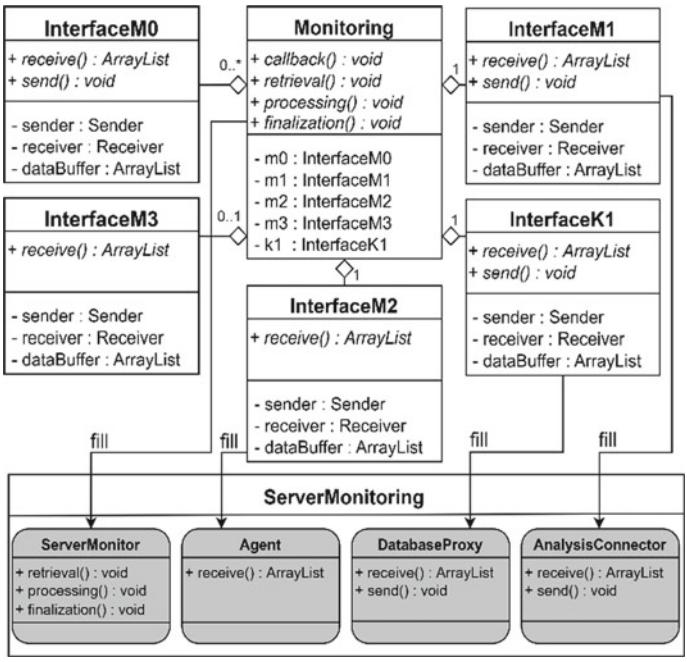



**Fig. 3** CROM model with all monitoring-associated naturals and one fictitious compartment (context) with roles for a subset of naturals

6. what *database system* is used, as the role that is played by the knowledge natural functions as a proxy between the system and one or multiple database management systems. Through having all necessary data saved within databases, the concept automatically supports *interfaces for external tooling to access all knowledge* from the system.

With that, the concept supports all criteria for flexibility from our classification in Sect. 2, except the criterion of a plugin engine. However, roles already act similar to plugins, thus, a dedicated plugin engine is not essential.

The flexibility of this concept comes with an upfront cost for planning and implementing the roles for a new use case, however, the cost of deployment and configuration of less flexible state-of-the-art systems is likely to be similar (subject for future qualitative evaluation). In the following, we want to discuss how such a system can be implemented.

### *3.3 Implementation of a Role-Based Monitoring System*

For the implementation of our concept, we used the programming model *Object-Teams Java* (OT/J), which originated from the issue that systems built from isolated objects suffer from increasingly poor structure, the more complex they get. In OT/J, this problem is lifted by natively allowing objects to play roles within contexts. The language definition of OT/J provides three main concepts, which are very similar to naturals, roles, and compartments from CROM:

1. *Base classes*, which are the classes to which a role is bound. *Base objects* (instances of a base class) have attributes and functions that describe their very nature, similar to naturals in CROM. Those attributes and functions may be inherited or overwritten by roles that are played during run time.
2. *Role classes*, whose instances bind to base objects by the *playedBy* keyword. Roles can declare *callout* and *callin bindings*. A callout binding allows a method call to a role object to be forwarded to its bound base method. A callin binding, on the other hand, allows a method call to a base object to be intercepted by its bound role method to be either executed *before*, *after*, or *instead* of the base method. In the context of an implementation of the concepts of this chapter, role classes of OT/J are identical to roles in CROM.
3. Finally, *team classes*, which are meant as containers for role classes. Every direct inner class of a team class is a role class. Teams encapsulate context in OT/J, similar to how compartments work in CROM. Teams can be activated, enabling callout and callin bindings from their contained roles to respective base objects.

To provide a frame for the implementation of our concept, we translate the CROM model from Fig. 3 into program code below (limited to the monitoring component).

The CROM primitives are easily translatable into OT/J. We start with the monitoring natural, for which we need to create an equivalent base class *Monitoring* (see

lines 1–12). The base class contains the definition of interface objects (line 2), a constructor (line 3) as well as the definition and implementation of the *callback()*, *retrieval()*, *processing()*, and *finalization()* methods (lines 4–11).

```
1   public class Monitoring {
2      InterfaceObject m0, m1, m2, m3, k1;
3      public MonitoringBase() { // Base Constructor }
4      protected void callback() { // Base Method
5        retrieval();
6        processing();
7        finalization();
8      }
9      protected void retrieval() {//Do nothing specific}
10    protected void processing() {//Do nothing specific}
11    protected void finalization() {//Do nothing specific}
12  }
13  public team class ServerMonitoring {
14    public ServerMonitoring(Monitoring as ServerMonitor sm) { //Constructor
15      activate(ALL_THREADS); // Without this, callin bindings have no effect.
16    }
17    protected class ServerMonitor playedBy Monitoring {
18      public ServerMonitor(Monitoring sm) { // Role Constructor }
19      callin void retrieval()  {//use-case specific retrieval}
20      callin void processing()  {//use-case specific processing}
21      callin void finalization()  {//use-case specific finalization}
22      // replace methods from base with methods from role.
23      retrieval <- replace retrieval;
24      processing <- replace processing;
25      finalization <- replace finalization;
27    }
28  }
```

Next, we translate the compartment of the CROM model into code by creating the equivalent team class *ServerMonitoring* (lines 13–28), that contains a constructor (lines 14–16) and the direct inner role class *ServerMonitor* (lines 17–27). The role class is bound to the *Monitoring* base class by using the *playedBy* operator (line 17). Furthermore, by using the *callin operator*, we define role methods (lines 19–21) that can intercept its bound method within the base class to be executed before, after, or instead of the bound method. By using the *replace binding* (lines 23–25), we define that the role method is used *instead* of the base method.

The exemplary modeling and implementation process that was shown in this section can be applied to all other control loop components and interfaces in the exact same way.

### 3.4 Applying Role-Based Monitoring to Different Application Domains

In the following, we will explain how this concept can be used to support the monitoring of systems in different application domains. The chosen application domains do not indicate any boundaries of the concept but were selected to highlight how the concept can be applied to common application domains for monitoring technology.

However, we want to emphasize that limitations for the concept's application exist. Due to its implementation with ObjectTeams, which is a Java based language, adding support for new application domains can only occur if the Java Virtual Machine (JVM) provides the functions and libraries that are needed to program the respective use-case specific roles. Nowadays, the landscape of libraries and the basic functions of the JVM support close to any use-case, however, this is no guarantee for support of all possible application domains.

We finalize this section with the application of our concept to the monitoring of *network devices*, the monitoring of *Ubuntu servers*, and the monitoring of *embedded systems*.

**Network Device Monitoring**

*Network devices*, also known as *network components*, are physical devices required for communication and interaction between devices in a computer network. In particular, they regulate the transfer of data in a computer network. Different active network devices (e.g., gateways, routers, bridges, access points, hubs, etc.) nowadays support common standardized interfaces for data communication and device management. For example, network device management is usually performed with SNMP. For that, SNMP allows the communication and modification of a variety of device parameters, including parameters that indicate network device performance such as applied load, memory utilization, bandwidth, and many more. Consequently, SNMP was used to monitor the performance of network devices for a long time. While SNMP also was successfully used in a control loop architecture for monitoring and managing servers [26], its main application domain remains the monitoring and management of network devices.

With our concept, adding support for monitoring and managing network devices through SNMP is an easy task. To achieve this, we need to create an implementation of roles that provide the basic communication primitives of SNMP. This includes implementations for querying parameters from the network device in the monitoring component over its M.2 interface (i.e., SNMPGet, SNMPGetBulk) as well as changing parameters of the network device in the execution component over its E.2 interface (i.e., SNMPSet). The analysis, planning, and knowledge components do not need to support SNMP specific primitives, yet their role-based implementation has to respect the data formats that are provided by the respective network device as those are usually carried through the whole control loop. In this case, the data representation format for network performance metrics is usually associated with either pure numerical values or with a SNMP Management Information Base (MIB) that

formally describes the SNMP objects and their provided values. While support for numerical data is trivial, the SNMP MIB is usually provided by the network device in question.

Support for the basic communication primitives of SNMP is not included in the JVM, thus, one needs to either include a third-party library (e.g., SNMP4J) that brings those functions or implement them from scratch.

Adding support for monitoring and managing devices that use SNMP with our concept shows that our concept supports niches that are not supported by most modern monitoring toolkits.

**Ubuntu Server Monitoring**

An application domain that is the main focus of monitoring systems is the monitoring of Ubuntu servers (and other Linux distributions). Due to this domain being widely supported, we can describe another strength of our concept, i.e., the use of existent third-party agents for monitoring.

Where other monitoring toolkits rely on using their own software agents (e.g., Elastic Stack uses Metricbeat, Prometheus uses Node Exporter), with our concept, the user can choose freely whether he wants to use an own software agent or software agents that are provided together with other monitoring toolkits.

Creating an own software agent is achieved through deploying a monitoring component natural on the system and creating a role-implementation that queries the needed parameters directly from the kernel over the M.2 interface via functions that are provided by the JVM. For using existent third-party software agents, the user needs to provide a role-implementation for the monitoring component that queries the interface of the agent over the M.2 interface (usually via HTTP/S) and filters for the needed data.

Due to the common support for this application domain, we use a scenario within it as our main evaluation scenario. This is discussed extensively in Sect. 4 of this chapter.

**Embedded System Monitoring**

Finally, we want to highlight another strength of our concept, which lies in creating support for monitoring embedded system product lines.

Embedded systems are nowadays included in close to every electrical device (e.g., washing machines, kitchen appliances, headphones, and many more). A large number of such devices executes Java programs to fulfil their purpose which enables the application of our concept for the monitoring and management of those devices.

By using our concept, it is easy to support whole product lines of similar devices as the concept separates different use-case specific implementations in roles under context. This allows manufacturers to maintain a generic monitoring and management toolkit based on our concept, while adding support for new appliances of the same product line through adjusting and implementing new roles. This results in less programming overhead, better maintainability and a common code base for different device iterations within the same product line.

As embedded devices are rarely equipped with powerful computation resources and large memory, we recommend to rely on baseline functions that are provided by the JVM for implementing role-based monitoring for embedded devices.

## 4 Evaluation Method

In this section, we will present our evaluation methodology that we use to support our claim that the role-based monitoring approach can provide a generally more flexible solution for monitoring IT systems while maintaining reasonable cost and performance. To this end, we first present our main evaluation scenario that is based on the Znn.com benchmark [5]. We also discuss three system architectures, one of which is based on the role-based approach and two others based on modern state-of-the-art monitoring toolkits. Finally, we describe how we will perform qualitative evaluation in the future and present results of an already executed quantitative evaluation.

### *4.1 The Znn.Com Benchmark for Self-adaptive Systems*

In the self-adaptive community, Znn.com [5] acts as a well-accepted proof of concept for applying self-adaptive techniques to classic web service scenarios. In the following, the Znn.com system will be introduced briefly to present the general idea behind its concept to the reader.

Znn.com is a system of servers and applications that mimics the infrastructure of actual news-provider services. It uses a three-tier architecture that consists of a set of web servers that serve static and dynamic content, a load balancer that distributes requests among the web servers (both of which run on Linux virtual machines), and stateless clients that access the web content.

During times with high amounts of accesses, usually during peak hours but also during unanticipated slash-dot[6] events, the managing subsystem adjusts the number of web servers to fit the resource demand. When the resource demand falls back to normal again, the number of web services is reduced again to fit the reduced resource demand. Also, Znn.com uses different policies regarding the content they serve to users of the web service. When experiencing a high load, the system adapts itself to present only content that uses a low amount of bandwidth (i.e., text). During times of adequate load, Znn.com also presents pictures and videos to its users. The execution of adaptations and data acquisition in Znn.com is handled by Perl scripts most of the time. Extracted data was saved within a MySQL database server.

The monitoring sensors that the authors developed to extract performance data were highly specific and tailored to this concrete example, matching their very own quality goals. The authors also stated that new probes would need to be developed for

[6] A sudden and heavy increase in performance demand.

other domains, without providing precise details on the tasks required for them. Thus, there is room for improvement regarding the flexibility of the monitoring methods used by Znn.com, as they require interface and sensor development for new use-cases to be supported. This is one reason why the concept of Znn.com is a good candidate for the application of our approach. Web services in general represent an important issue in today's information technology, which also highlights the need to explore the benefits that a flexible role-based monitoring solution can provide to this landscape compared to traditional monitoring techniques.

## *4.2 Auto-scaling of Web Service Infrastructures*

Below, three versions of an own and adjusted implementation of the Znn.com concept will be discussed, that each use different technologies, allowing to compare the role-based monitoring approach with modern state-of-the-art monitoring toolkits. All three versions only use the general idea behind Znn.com but are re-designed from scratch.

1. First, a role-based version as described in the previous section based on Object-Teams Java for managing and Docker Swarm[7] for providing orchestration capabilities.
2. Second, to compare the role-based approach to state-of-the-art infrastructures, the Prometheus monitoring and alerting system is used. Again, a Docker Swarm environment is used for orchestration.
3. Finally, to have yet another comparable state-of-the-art system, Elastic Search[8] and Kibana[9] are used within a hybrid Kubernetes[10] (for running the Elastic Stack) and Docker Swarm (for orchestrating the web service) environment as the third version.

In all three versions, the infrastructure is matched as closely as possible to the abstract MAPE-K pattern [31], which is depicted in Fig. 4. It consists of one load balancer and an arbitrary number of servers for distributing the load. The node that contains the load balancer also contains the managing subsystem, whereas the nodes that contain the web servers act as the monitored systems that are monitored and managed by the managing subsystem. The load balancer possesses four MAPE-K components (MAP-K), whereas each of the web servers possesses only one monitoring and one execution component.

As noted earlier, this chapter's version of the web service implements some adjustments to the core functionality of Znn.com:

[7] Docker https://www.docker.com/ and Docker Swarm https://docs.docker.com/engine/swarm/.

[8] Elastic Stack https://www.elastic.co/de/elastic-stack.

[9] Kibana Website - https://www.elastic.co/de/kibana.

[10] Kubernetes Website - https://kubernetes.io/

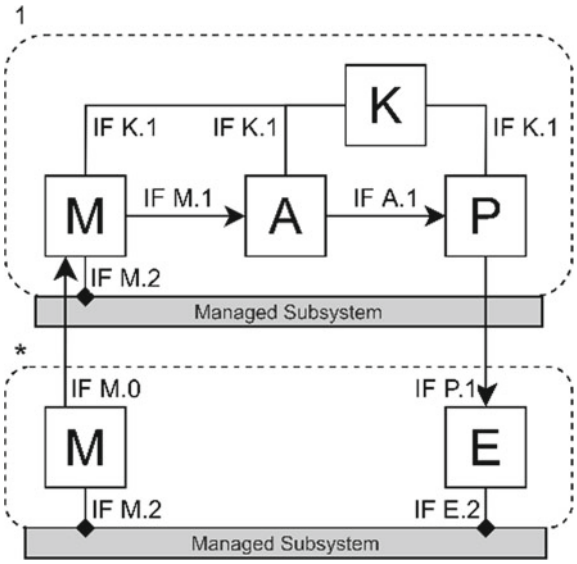



**Fig. 4** MAPE-K pattern [31] of the Znn.com-inspired architecture for evaluation

1. As the idea of showing different types of media in times of high load almost always induces loss of information, the adjusted implementation will provide the same type of web service to users regardless of load status.
2. As a consequence of 1, the adjusted implementation does not need to change the content fidelity in the same way as Znn.com did. Therefore, the only adaptation that is executed in the adjusted implementation is the increase and decrease in servers that run a replica of the web service.
3. Znn.com mimicked actual news provider web services by presenting text, videos, and images to arbitrary topics. While this is representative of a real-world scenario, it brings along the disadvantage of not allowing to adjust the load created per user visit easily. This, however, is necessary to provide an extensive quantitative evaluation of this example. To address this, the adjusted implementation relies on using a mathematical operation (i.e., calculating the factorial of a large number) for each visit rather than presenting actual content to virtual visitors.

To make the three versions comparable, an expected set of functions for each of the components is explained below.

- The monitoring component on each web server (the agent) has a direct connection to the sensors on the underlying operating system of the server. Its sole purpose is to aggregate data about the respective server through the M.2 interface and forward this data to the main monitoring component on the load balancer through the M.0 interface. The monitoring component on the load balancer aggregates the data from all agents, and itself gathers data from the sensors on the load balancer.

All data is organized as a time series and transported through the K.1 and M.1 interfaces to the knowledge- and analysis components, respectively.

- The analysis component has a direct link to the monitoring component (IF M.1). Periodically, the main monitoring component sends a set of metrics through the M.1 interface that represents the system's state. This set is analyzed, whether the performance of the infrastructure is sufficient or if the load applied to the infrastructure has surpassed a given threshold. This information is passed over the A.1 interface to the planning component.
- In this example, the planning component must choose between two possible actions, based on the provided decision from the analysis component. That is, firstly and secondly, the increase or decrease of the number of web servers that are accessed by the user through the load balancing configuration. As the adaptation decision contains data about which metric was decisive in the first place, it implicitly infers the appropriate action to execute. Based on this action, a notification is created and passed over the P.1 interface to the execution component (or the external API in versions two and three of the scenario).
- How those actions are executed depends on the technology that is used to set up the infrastructure. As was shown in Sect. 2 on related work, most state-of-the-art monitoring systems do not allow to natively execute actions on local or remote machines, thus, they need to rely on a specifically tailored external web API. The role-based approach does not have this limitation. How the different versions of this evaluation example handle execution will be fully discussed in a future publication. The requirement for each version, however, is to trigger the increase or decrease of the number of web servers that are accessible for users.
- The knowledge component keeps track of all data from the monitoring components and provides an interface to the database of the system embodiment.

### *4.3 Qualitative and Quantitative Evaluation*

Finally, we want to explain how we want to conduct a qualitative evaluation in the future. This is accompanied by a presentation of the results on a quantitative evaluation of the role-based monitoring approach that was already executed.

**Qualitative Evaluation**

During our qualitative evaluation, which will be the subject of a future publication, we want to discuss different metrics that cannot be expressed as clear-cut numbers. During the evaluation, first, we will discuss how the three versions of the infrastructure approach flexibility and how flexible their functions are provided to the user.

Furthermore, we discuss the entry barriers that are involved for each of the versions, including programming languages, infrastructure technologies such as virtual machines and automated provisioning, operating system knowledge, management of orchestration environments, and more.

Next, we will discuss the overhead that is induced by using either of the three versions. For this, we will give estimates on how many hours the implementation consumes as well as how much effort is needed to apply those infrastructures to other use cases. Finally, we rank the usability of those systems from a functional perspective.

**Quantitative Evaluation**

During our quantitative evaluation, we performed three main experiments. In the first experiment, we collected evidence of the correct functionality of all three infrastructures. In the second and third experiments, we collected metrics that allowed us to draw conclusions about the performance of the role-based monitoring approach compared to modern monitoring toolkits.

During all experiments, we measured the CPU load on all participating machines. If a machine running a web service exceeds an average CPU load of 30% over a period of more than 10 s, all versions generate an alert that sends a notification to the execution environment and causes a new machine to be started (if auto-scaling is enabled). If the CPU load drops below an average value of 20% for a period of more than 10 s, the same procedure is performed, but one machine is removed (if auto-scaling is enabled). During the experiments, the load on the infrastructure is gradually increased by sending an increasing number of HTTP requests to the web servers through a load balancer. Each request causes the web service to compute the factorial of a large number (35,000!), creating a fixed load each time.

*First Experiment*

In the first experiment, we run two iterations of a load test on all three versions of the infrastructure. During the load test, we start with one request per second and double the number of RPS every minute until we reach a defined maximum. After the maximum is held for four minutes, the number of RPS is halved every minute until the experiment ends. In the first run, auto-scaling is disabled for the infrastructure. In the second run, auto-scaling is enabled.

We expect the infrastructure to function without failure in both cases. However, without adaptation enabled the load measured on the involved web server will be significantly higher compared to the run with adaptation enabled. With adaptation enabled, we expect the infrastructure to successfully shift load to additional web servers, as the number of machines involved will automatically increase in the second run. We further expect that, in both iterations, the load on the involved management node that contains the adaptation logic and the load balancer should remain similar.

Table 2 shows the results of the first experiment which meet our expectations. While the load on the management nodes in all three infrastructures remains similar between the two iterations, we see significant decrease in CPU load on worker nodes after enabling auto-scaling (adaptation).

This is also reflected by Figs. 5 and 6 which show the load curve for active worker nodes for the role-based monitoring approach (version 1) in both iterations of the experiment.

Here, we can see significant changes in the shape of the load curve in the respective graphs. Without adaptation, the load curve rises as soon as more load is applied on the

**Table 2** Comparison of the average CPU usage of Elastic Stack, Prometheus, and the role-based monitoring approach (RBMA) without adaptation (top), with adaptation (middle), and their numerical differences (bottom)

| Metric | Elastic Stack (%) | Prometheus (%) | RBMA (%) |
|---|---|---|---|
| Avg. CPU% (Manager) | 8 | 4 | 2 |
| Avg. CPU% (Worker) | 65 | 64 | 66 |
| Peak CPU% (Manager) | 36 | 13 | 11 |
| Peak CPU% (Worker) | 97 | 100 | 99 |
| Avg. CPU% (Manager) adapted | 8 | 5 | 3 |
| Avg. CPU% (Workers) adapted | 35 | 49 | 34 |
| Peak CPU% (Manager) adapted | 30 | 13 | 9 |
| Peak CPU% (Workers) adapted | 74 | 97 | 98 |
| Avg. CPU% (Manager) diff | +0 | +1 | +1 |
| Avg. CPU% (Worker) diff | −30 | −15 | −32 |
| Peak CPU% (Manager) diff | −6 | +0 | −2 |
| Peak CPU% (Worker) diff | −23 | −3 | −1 |

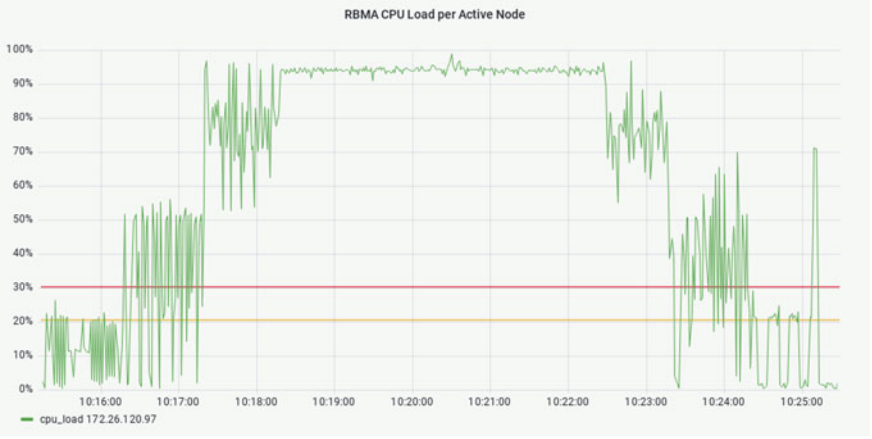



**Fig. 5** Representation of the worker node CPU% load (RBMA) without adaptation enabled

infrastructure, capping out at 98% (version 1). With adaptation enabled, we can see that this load is now offloaded onto two more nodes that host the same web service, reducing the average load by 32%.

In both iterations, we can recognize high peak loads. Without adaptation, those peak loads are caused by general high load on the involved worker node. With adaptation enabled, this peak load is caused by initial spikes of high load whenever a new machine is added, as depicted in Fig. 6. While the peak load numbers are




**Fig. 6** Representation of the worker nodes CPU% load (RBMA) with adaptation enabled

similar between both iterations, the load curves show that enabling adaptation causes an overall healthier peak and average performance behavior.

*Second Experiment*

In the second experiment, we measure the average CPU load on all participating nodes in all three versions of the architecture during a 60-min idle period. Table 3 shows four metrics that indicate the idle performance of the three infrastructures.

From the results, we can derive that all three infrastructures consume a similar amount of CPU resources on average during idle with RBMA consuming the smallest amount and Elastic Stack consuming the largest amount. Regarding CPU load peaks, Prometheus produces the smallest peak loads while Elastic Stack produces the highest peak loads.

*Third Experiment*

For the third experiment, we ran a total of 10 iterations of the same load test that was already used in experiment one, however, this time we focus on measuring long-term performance. Thus, we use additional metrics for evaluating the three infrastructures. All metrics are calculated as an average over all 10 iterations of the experiment and are listed in Table 4.

**Table 3** Comparison of the average idle CPU usage of Elastic Stack, Prometheus, and the role-based monitoring approach (RBMA)

| Metric | Elastic Stack (%) | Prometheus (%) | RBMA (%) |
|---|---|---|---|
| Avg. CPU% idle (Manager) | 7 | 4 | 3 |
| Avg. CPU% idle (Workers) | 2 | 3 | 2 |
| Peak CPU% idle (Manager) | 36 | 18 | 29 |
| Peak CPU% idle (Workers) | 3 | 3 | 1 |

**Table 4** Comparison of the average performance of Elastic Stack, Prometheus, and the role-based monitoring approach (RBMA) over ten experiment iterations

| Metric | Elastic Stack | Prometheus | RBMA |
|---|---|---|---|
| Avg. CPU% (Manager) | 9% | 4% | 3% |
| Avg. CPU% (Workers) | 39% | 52% | 39% |
| Peak CPU% (Manager) | 35% | 18% | 13% |
| Peak CPU% (Workers) | 82% | 97% | 98% |
| Avg. first > 30% to scale-up | 31,1 s | 110,1 s | 27,8 s |
| Avg. Alert to new node avail | 5,3 s | 4,2 s | 6,7 s |
| Avg. Alert to node removed | 3,5 s | 4,6 s | 5,5 s |
| Avg. scaling operations | 10 | 4 | 7,6 |

During the experiments, the working nodes running Prometheus generated the highest load on average, followed by Elastic Stack and RBMA, which both generated the same load.

Load peaks were highest on RBMA, followed closely by Prometheus. Elastic Stack generated the lowest load peaks during the experiment. Again, with adaptation enabled, high peak loads are a result of the booting process of new nodes, as reflected by Fig. 6. As the load measurements begin as soon as a new machine is online, usually only the first one or two measurements are highly elevated.

Next, we measured the average time between the first load measurement exceeding the 30% threshold and the availability of a new working node. RBMA performed best on this metric, followed closely by Elastic Stack. Prometheus took more than three times as long on average to provision a new node. This is because all three versions were configured to only raise an alert if the threshold was exceeded for more than 10 s. If the threshold of 30% was not surpassed once during this time, the 10 s period was reset. The results that are indicated by this metric do not tell that Prometheus is far worse in recognizing high load, however, they tell that with the exact same configuration, Prometheus scales less aggressive than Elastic Stack and RBMA. Prometheus could be configured to scale as aggressively by reducing the overshoot protection interval from 10 s to a lower time interval.

Next, we measured the time between when an alert was triggered and when a new working node became available. Prometheus performed best, with an average time to scale of 4.2 s, while Elastic Stack took 5.3 s and RMBA took 6.7 s. We also measured the time to scale down, with Elastic Stack performing best, followed by Prometheus and then RBMA.

Finally, we measured the average number of scaling operations (upscaling and downscaling) during the experiments. The Elastic Stack scaled more frequently (10 times per iteration) than RBMA (7.6 times) and Prometheus (4 times).

In summary, the performance of all three versions was close. All three versions were able to scale appropriately to offload an unsustainable load from one machine by dividing it among a maximum of three machines. This shows that RBMA can

provide a more general and flexible monitoring solution while maintaining reasonable performance.

## 5 Outlook and Future Work

Within this chapter, we have provided an overview of our currently ongoing research on a flexible role-based context-aware monitoring system with a control loop structure. We have provided an overview of different techniques for developing flexible systems and have shown how other monitoring systems approach flexibility. We discussed a comparison of fifteen monitoring systems from different decades and evaluated them against eleven flexibility criteria. We then discussed the notion of roles and how it can be applied to software to increase flexibility. We introduced a concept based on role-based control loop components that can work in different application domains. We presented an associated model and outlined the implementation of such a concept as well as how this concept can be applied in different application domains. In parallel, we discussed a scenario and system structure for evaluating this concept and presented our method for evaluation. Finally, we have discussed our plans to perform a qualitative evaluation in the future as well as the results of an already conducted quantitative evaluation.

Going forward, we plan to take two additional steps to complete this research project. First, we plan to actually conduct our qualitative evaluation in the near future. Next, we want to apply the role-based monitoring approach to another application domain to gain further insight into the flexibility of this approach.

In addition, this research project opens up another area of research, namely the application of a generic role-based control loop to domains other than information technology systems monitoring.

## References

1. Bachman CW, Daya M (19977) The role concept in data models
2. Bae TM (2014) System for measuring transmission bandwidth for media streaming and method for same. USA, Patentnr. US Patent 8,625,443
3. Bäumer D, Riehle D (1997) Product trader. Citeseer
4. Bäumer D, Riehle D, Siberski W, Wulf M (1998) The role object pattern. Washington University Dept. of Computer Science: Citeseer
5. Cheng S-W, Garlan D, Schmerl B (2009) Evaluating the effectiveness of the rainbow self-adaptive system
6. Christiansen J, D'angona R, Bell C (2014) Method and system for assessing, managing, and monitoring information technology risk. USA, Patentnr. US Patent 8,744,894.
7. Dobson S, et al (2006) A survey of autonomic communications. ACM New York, NY, USA
8. Freeman E, Robson E, Bates B, Sierra K (2008) Head first design patterns. O'Reilly Media, Inc., Newton
9. Gamma E, Helm R, Johnson R, Vlissides J (1995) Elements of reusable object-oriented software. Addison-Wesley Reading, Boston

10. Guarino N (1992) Concepts, attributes and arbitrary relations: Some linguistic and ontological criteria for structuring knowledge bases
11. Harrow JR, Messinger FP (1994) System monitoring method and device including a graphical user interface to view and manipulate system information. USA, Patentnr. US Patent 5,375,199
12. IBM (2006) An architectural blueprint for autonomic computing. IBM Corporation
13. Ingalsbe D, et al (2007) Monitoring Technology. USA, Patentnr. US Patent 7304588B2
14. Johnson RE, Zweig JM (1991) Delegation in C++. J Object-Oriented Program 4:31–34
15. Kephart JO, Chess DM (2003) The vision of autonomic computing
16. Kini V (1999) Computer server with improved reliability, availability and serviceability. USA, Patentnr. US Patent 5,864,659
17. Krupitzer C, et al (2018) A survey on engineering approaches for self-adaptive systems (extended version)
18. Kühn T, Böhme S, Götz S, Aßmann U (2015) A combined formal model for relational context-dependent roles, New York, NY, USA
19. Maiocco JN, Belanger BM, Mansell BEMccaskey JA (2012) System and interface for monitoring information technology assets. USA, Patentnr. US Patent 8,239,527
20. McCrory DJ, Gold AP, Sanderson A (2004) Remote computer system monitoring and diagnostic board. USA, Patentnr. US Patent 6,697,962
21. Nasle A, Nasle A, Meagher K (2010) Systems and methods for real-time system monitoring and predictive analysis. USA, Patentnr. US Patent 7,826,990
22. Pereira AA, Javed MA, Lam CH, Yuen C-Y (2010) Method and apparatus for system monitoring and maintenance. USA, Patentnr. US Patent 7,702,435
23. Salehie M, Tahvildari L (2009) Self-adaptive software: landscape and research challenges. ACM, New York
24. Shmelkin I (2020) Monitoring for control in role-oriented self-adaptive systems
25. Shmelkin I, Schütze L, Kluge T (2020) Modeling flexible monitoring systems with a role-based control loop, Berlin
26. Shmelkin I, Springer T (2021) On adapting SNMP as communication protocol in distributed control loops for self-adaptive systems
27. Sowa JF (1984) Conceptual structures: information processing in mind and machine
28. Steimann F (2000) On the representation of roles in object-oriented and conceptual modelling
29. Vinberg AB, et al (2010) Model-based system monitoring. USA, Patentnr. US Patent 7,802,144
30. Weyns D (2020) An introduction to self-adaptive systems: a contemporary software engineering perspective. John Wiley & Sons, Hoboken
31. Weyns D, et al (2013) On patterns for decentralized control in self-adaptive systems. In: Software engineering for self-adaptive systems II: international seminar, Dagstuhl Castle, Germany, 24–29 October 2010 Revised Selected and Invited Papers. Springer, Heidelberg, pp 76–107
32. Wolters TJ (2006) Predictive monitoring and problem identification in an information technology (IT) infrastructure. USA, Patentnr. US Patent 7,107,339

# Adaptation Consistency of Distributed Role-Oriented Applications Based on the Actor Model of Computation




**Daniel Matusek, Tim Kluge, Ilja Shmelkin, Thomas Springer, and Alexander Schill**

**Abstract** Complex software is expected to adapt to changing circumstances autonomously. In software engineering, structural adaptations of objects can be modeled using contexts and roles. Existing role concepts support autonomous adaptation but provide only limited support for distributed setups. This paper addresses two major issues of distributed adaptive systems. Firstly, the mapping of roles to the actor model, which facilitates the concurrent execution of independent objects and their communication via message passing. Secondly, a concept for coordinating adaptations across multiple distributed nodes, relying on a strong or eventual consistent adaptation of distributed role-based applications. The former approach always ensures a globally consistent state of the system. The latter increases the system's availability in the case of network congestions or failures of devices during the adaptation process by superseding a rollback of all intermediate changes. We discuss the cases for applying different consistency criteria and show the feasibility of our approach by presenting an empirical case study of a file server application in a prototypical programming language.

**Keywords** Self-adaptive · Role concept · Actor model · Distributed systems · Consistency

D. Matusek (✉) · T. Kluge (✉) · I. Shmelkin (✉) · T. Springer (✉) · A. Schill (✉)
Faculty of Computer Science, Technical University Dresden, Dresden, Germany
e-mail: daniel.matusek@tu-dresden.de

T. Kluge
e-mail: tim.kluge1@tu-dresden.de

I. Shmelkin
e-mail: ilja.shmelkin@tu-dresden.de

T. Springer
e-mail: thomas.springer@tu-dresden.de

A. Schill
e-mail: alexander.schill@tu-dresden.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_2

## 1 Introduction

Modern computer systems are increasingly interconnected and form different types of networks. That leads to the increased complexity of the overall system structure. Self-Adaptive software systems (SAS) aim at dealing with the complexity of maintaining those systems autonomously and can modify their behavior and internal structure according to environmental changes. Intuitively, highly distributed systems from the Internet of Things domain, autonomous cars, or smart applications which need to adapt their behavior autonomously as a response to their environment can be modeled as SAS to supersede human interaction regarding the adaptation process.

An essential aspect of self-adaptation is executing changes during run-time without shutting down nodes or the overall application. A possibility to accomplish that task is the usage of roles [13]. Roles enable the switch from one system behavior to another during run-time, capturing the context and providing behavior for the specific context. Additionally, the role concept can be used to model both the static part of the application and the dynamic part, which can be adapted. Nevertheless, current solutions for role runtimes like OT/J, SCROLL, and LyRT [16–18] lack a natural approach for distribution and concurrency [17, 18] which would be beneficial regarding distributed and decentralized adaptive applications.

Therefore, this work presents an approach for concurrency and distribution for roles using the actor model of computation with a novel DSL called *Rolactor*, which models software systems as a set of concurrent entities communicating via message passing [15]. By integrating the actor model of computation with role-based programming, we can combine the advantages of both approaches. By introducing a DSL that abstracts the adaptivity, concurrency, and distribution of the application, the complexity of these aspects can be managed by the system and is hidden from the developer. Executing adaptations in distributed systems requires the means to coordinate an adaptation involving multiple participants. In our DSL, we support defining transactions that can be used to coordinate adaptations of multiple actors using roles. We adopt a transaction protocol, allowing for consistent adaptation without data loss by applying the approach for decentralized adaptation execution [10]. Transactional approaches require all actors to execute those changes atomically without any failures.

Furthermore, those actors must wait for others to activate those changes transactionally. In large systems, where multiple nodes can be adapted, this strict approach can slow down the whole adaptation process or prevent already adapted nodes from working if a failure occurs that would not affect the functionality. Therefore, we investigate how to relax the strict transactional adaptation process to increase the availability of single actors during run-time since no sophisticated approach to eventual consistent self-adaptation execution exists to the best of our knowledge.

This work is outlined as follows. Chapter two introduces the case study of an adaptive distributed file server application to show the applicability and feasibility of our concept. Chapter three presents our work's basic concepts and the necessary foundations for better understanding. In chapter four, we present our idea of

a distributed role runtime using the actor model of computation to deal with the distribution of roles and concurrency, considering different types of consistency to allow for consistent transitions from one configuration to another in case of an adaptation. We evaluate the concept by applying it to our case study to show its necessity and applicability in chapter five. In chapter six, we present related approaches and discuss their limitations regarding our requirements. Last, chapter seven summarizes the paper and gives an outlook on future work.

## 2 Case Study

To evaluate the concept presented later in this work, we introduce a case study of a distributed file system motivated by Taing et al. [1], which we extended with additional functionality. Multiple clients can download large files from two file servers, where the files are split into smaller chunks because of their size. The transfer process is subject to adaptation because single clients can request the transfer to be encrypted. In that case, the server must change its role to satisfy that demand as well as the requester to ensure that the received data gets decrypted correctly. This encryption behavior can be enabled during run-time without stopping the application and restarting the server.

The file transfer application not only supports the encryption or decryption of the data, but the clients also consider their connection strength regarding the data transfer from the server to the client. Clients can require the server to send images either compressed or decompressed. In that case, a client with a weak connection issues adaptation operations to the servers and requests them to play the "compressed" role. In that case, the servers should switch roles and send the compressed data to the peer.

Changing behavior is directly affecting the current role configuration. Ordinarily, the clients and the server play the "plain text" role, where files are transferred unencrypted from node to node, as depicted in Fig. 1. In addition, the servers play the role of "uncompressed" and send the files in their best quality. If any of the clients request an encrypted transfer, the change must be applied to the transmitting server of the file transfer system. Thad adaptation leads to potential issues regarding the consistency of the change, which will be discussed in the following section. The term consistency in our work refers to the overall system configuration of all nodes. We do not refer to the internal consistency of data or the transmitted data between the peers.

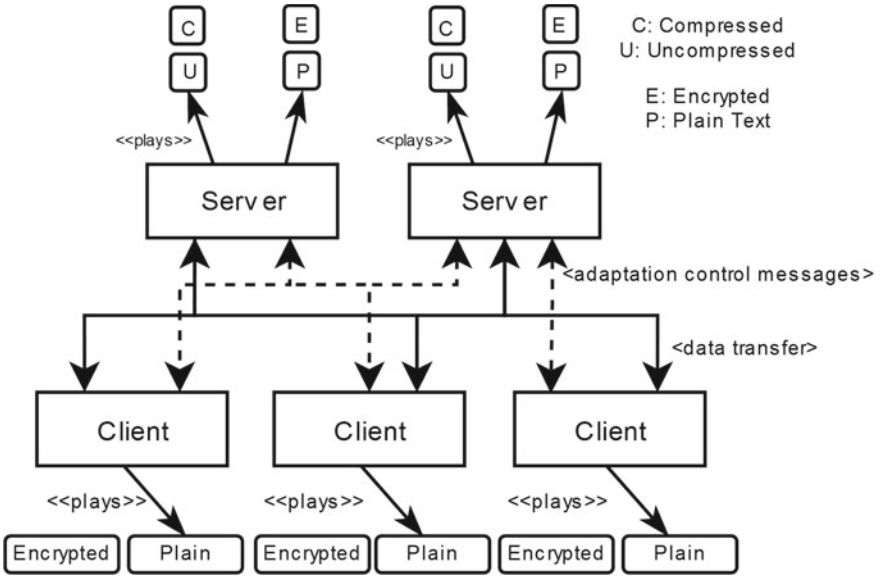



**Fig. 1** Structure of the extended file transfer application and their roles

## *2.1 Error Model*

Changing the current configuration can lead to consistency problems regarding the system's configuration. Suppose the whole system from Fig. 1 wants to enable encryption, but one of the clients fails to do so. In that case, the overall consistency of the configuration is violated, leading to erroneous data received from the server at the failed client. The failing client wants to read encrypted data but cannot decrypt it because the wrong role is played.

This inconsistency can occur because of two different cases. Firstly, an adaptation might fail for one of the nodes, which leads to an inconsistent global state because not every participant was adapted successfully. Secondly, as depicted in Fig. 2, an inconsistency can also be caused by the run-time adaptation of running processes. A switch from the unencrypted role to the encryption is issued. The dotted line indicates the point in time of the adaptation. The white boxes before the adaptation process non-encrypted data, whereas the blue boxes represent the new behavior, the encryption. Client 1 (c1) is still downloading data from the server, and c2 requests the transfer to be encrypted. The change does affect not only c2 but also the ongoing process of c1, which results in inconsistent received data because the chunks of a large file are unexpectedly encrypted, and c1 did not adapt to the change.

The presented case study can be divided into two parts. Firstly, the adaptation from plain text data transmission to encrypted behavior should be executed consistently to ensure that every node can read the received data. Figure 2 illustrates how crucial a consistent adaptation in such a use case is. The second adaptation between

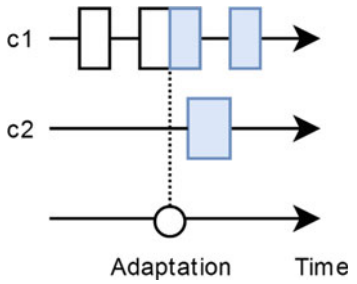



**Fig. 2** Possible inconsistency during the execution of adaptations

a compressed and decompressed data transmission is used to show the feasibility and applicability of our relaxed adaptation execution. If any server sends a full-sized image instead of a compressed one because the adaptation failed, the transmission time increases, but the integrity of the overall system is still preserved. Details regarding the use case are discussed in Sect. 5.

## 3 Foundations

In this section, we explain the foundations for the presented concepts. First, we explain the basics of SAS, a basic concept for developing our autonomous system. SAS need to be enriched by further concepts, mainly from the software engineering and distributed systems domains, to form usable frameworks. Therefore, we present the role concept as the foundation for run-time adaptations because of their contextual, behavioral, and relational nature and intuitive support for dynamic change. Hence, our DSL relies on roles to support the intuitive adaptability of applications. Afterward, we introduce adaptation operations and transactions, which are key structures for consistent adaptation in a distributed setting and are therefore needed to ensure that the actors remain in a valid configuration throughout the run-time. Next, the MAPE-K adaptation feedback loop is introduced. It is a concept for the distribution of the runtime elements and can be seen as an extensible blueprint for architecting self-adaptive systems. Furthermore, we explain the basics of the actor model of computation with its natural support for concurrency and distribution, which is beneficial for performing adaptations in a decentralized setting.

### *3.1 Self-Adaptive Software Systems*

Self-adaptive Software Systems (SAS) aim to tackle the growing complexity of increasingly distributed and possibly decentralized software systems. SAS supersede the need for human interaction to perform updates or adapt to changes since devices

can adapt autonomously by realizing objectives pre-defined by administrators or using machine-learning techniques to adapt to unknown circumstances. To achieve that, a SAS monitors its context and internal state and derives a plan on how to adapt if the current circumstances change and if an adaptation is necessary. This helps reduce the system complexity and human time spent maintaining running systems.

According to the literature, we can distinguish between internal and external adaptation [5]. The internal self-adaptation handles uncertainties and changes in the operational environment within its source code. It is hard to detect external influences and an internal approach hinders getting an overview of the overall system. Moreover, the complexity of the application's source code increases drastically by extending it with adaptation mechanisms. We distinguish between parameter adaptation, where specific values or parameters of the application are modified, or structural adaptations, where the architecture and overall structure are adapted.

Contrary to internal adaptation, external self-adaptation can be used. The adaptation logic gets separated into a managing and a managed part. By using that approach, we can better integrate external influences. In addition, depending on the implementation, a managing layer could possibly be used in different domains without a re-implementation. This approach is often realized as a closed-loop mechanism, such as MAPE-K, which is explained in Sect. 3.4.

## *3.2 The Role Concept*

With the notion of roles, we can express context-dependent behavior and change the structure of an application during run-time. Due to the context dependence, we can already model those context switches before running the system. Bachmann et al. [14] first mentioned roles in the literature in the context of data modeling. Several works refined the role concept in the following years, but no standard definition of roles and their structures exists, leading to terminology inconsistencies. Nevertheless, most approaches have players playing one or more roles in the system. Roles have three different natures: contextual, behavioral, and relational. The first applies to systems where a particular role is played for a specific scope of the application. In our example, a server can play the role "unencrypted" in a private environment and send data in plain text without any encryption.

In contrast, "plain text" can be activated, e.g., in a public network. Depending on the current context, different behavior is active for the same player (the server). The relational nature of roles highlights their collaboration aspect. Upon collaborating with other entities, e.g., a server and client relationship, the roles depend on the role they are collaborating with since one cannot exist without the other. With this abstraction, we can express the responsibilities of the single players according to their played roles. Finally, roles introduce behavior to their players, e. g. by introducing new methods to the base object.

Steimann [12] defined a feature set for roles at design- and run-time, which was extended by Kühn [13]. A role-based application can fulfill those features but does not

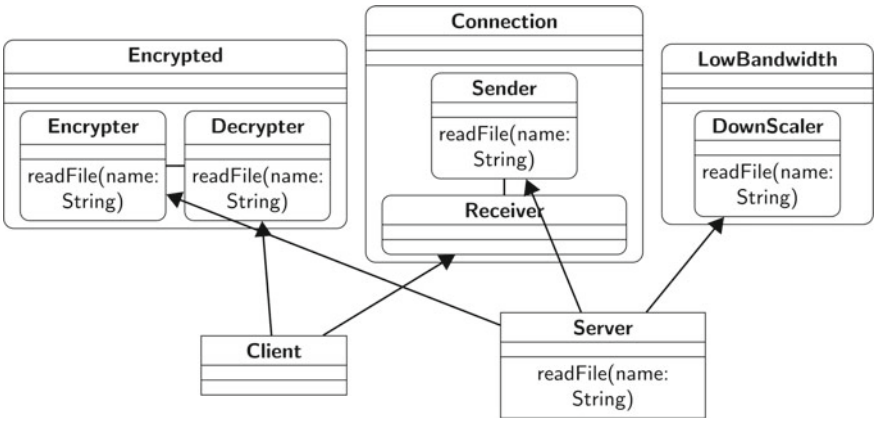



**Fig. 3** CROM diagram of the case study example

necessarily need to. Kühn [13] introduced the concept of "Compartments." Compartments depict objectified hierarchical context and can be used to model different contexts so that a player can play different roles in separate compartments. The "urban" environment or the highway could be modeled as a compartment where specific roles are played. The CROM meta-model defines the relationships that roles enter with each other and constraints between individual roles: Roles may imply that their player plays another role type simultaneously or prohibit the player from playing it.

With its context-dependent nature and resulting adaptability, the role-based approach can be used as the foundation for self-adaptive software systems. We aim for a run-time reconfiguration of the system depending on the current context.

In Fig. 3, an exemplary CROM diagram for our file server case study example is depicted. Here, *Client* and *Server* are represented as natural objects. Both can enter a relationship with each other by playing a *Sender* and *Receiver* role in a connection compartment. Adaptations can be triggered by creating an *Encrypted* compartment where client and server play their respective roles, or by introducing a *DownScaler* role that compresses requested images. The relationships between roles in compartments are depicted as a simple line.

## *3.3 Adaptation Transactions and Operations*

The concept of transactions is well known in the database domain. Transactions ensure that data is written consistently on multiple databases in a distributed setting. Motivated by that approach, Weißbach et al. [10] introduced the terms "Adaptation Operation" and "Adaptation Transaction," which have a similar nature compared to

database transactions. An adaptation operation is a step inside an adaptation transaction to modify the current role configuration. By using an operation, either binding between roles and players inside a role-based application or a collaboration between roles can be modified. An operation contains the source configuration and the target configuration, respectively. The source configuration identifies the subject to be adapted, such as the player and the roles, whereas the target configuration denotes the final state of the performed adaptation. Adaptation operations are grouped into adaptation transactions because they need to be executed together to keep the system in a consistent state and ensure that no data gets lost. Adaptation transactions are executed in an atomic way, i.e., all operations in a transaction succeed, or the system remains in its old state. With that approach, Weißbach et al. [10] aimed to achieve overall consistency during the system's run-time.

### *3.4 The MAPE-K Feedback Loop*

The MAPE-K (Monitor, Analyze, Plan, Execute, Knowledge) Feedback loop is crucial for performing adaptations in self-adaptive software systems. Kephart and Chess [2] present the concept of a so-called autonomic element, which is depicted in Fig. 4. It consists of one or more managed elements, i.e., a managed application (the role runtime) and the adaptation management responsible for the adaptation process. The management is performing adaptations on the application during run-time.

Generally, the adaptation management (AM) first monitors the internal state of the managed application and its external environment (see Fig. 4). Next, the environmental data and state is analyzed, and if required, a plan will be generated by the planning component. The execution stage of the MAPE-K feedback loop is

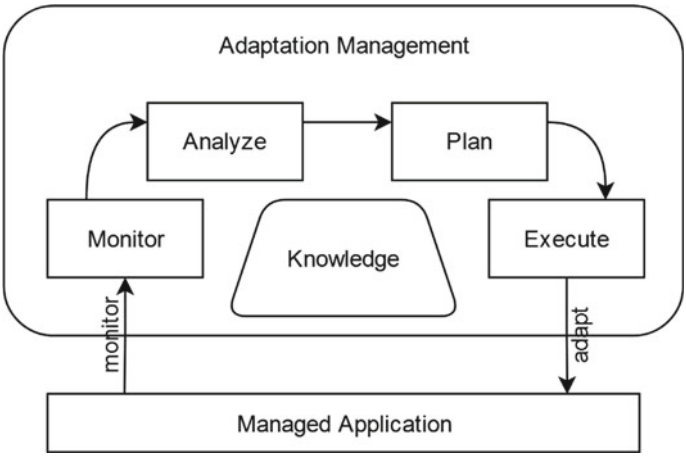



**Fig. 4** Autonomic element in self-adaptive software systems according to [2]

executing the planned change on the managed application to adapt to the changed circumstances. The knowledge component serves as a repository for environmental data or historical data for the single entities of the feedback loop to support the self-adaptation process.

The AM and the managed application can be distributed in various ways depending on the current needs of the maintained application. Weyns et al. [3] discuss different decentralization patterns for distributing the single elements of the MAPE-K feedback loop with their advantages and drawbacks. In our previous work [4] about intent-based adaptation coordination in decentralized networks, we extended the work from Weyns et al. [3] by introducing a hybrid decentralization pattern for a master–slave network architecture. This hybrid approach allows a centralized master node to make decisions for a large IoT cluster. In contrast, the slave devices gather monitoring data from their perspective to support the decision-making process of the analysis and planning component of the master MAPE-K loop.

### *3.5 The Actor Model*

The actor model was initially proposed by Hewitt et al. as a reference architecture to develop artificial intelligence software [3]. Actor systems are composed of a hierarchy of actors, which are entities executed concurrently with each other. Actors interact by asynchronously sending messages to each other. Each actor places incoming messages in a message queue and sequentially processes them by executing any number of the following statements per message:

- Sending a message to another known actor
- Creating a new child actor
- Replacing the message handling behavior

Thus, actor systems are highly concurrent by design and fully asynchronous. Actors can be transparently distributed across physical devices on a network, as messages can be transmitted.

Over the years, the actor model has been extended and adapted to facilitate the requirements of modern systems. As state-of-the-art systems are commonly developed following the object-oriented development paradigm, objectified actors have been proposed in the form of active objects [21]. Active objects are actors that resemble the shape of objects, i.e., they implement a class that defines a set of methods and fields. Compared to standard object-oriented programming, methods are always invoked asynchronously via message passing as in actor-based systems. Fields are never exposed to other active objects. Asynchronous message exchanges following a request-response pattern can be expressed using the abstraction of futures, which represent values that will be available at some point in the future when a message by the corresponding active object is received [21].

# 4 Concept Architecture

In the following chapter, we present our concept for a novel DSL called *Rolactor* that combines the actor model of computation with role-based programming, thereby facilitating the development of self-adaptive distributed systems with roles. By using the DSL, the case study program introduced in Sect. 2 using role-playing actors is enabled. Subsequently, we extend the language and its runtime to enable consistent adaptation using transactions and relaxed adaptation for eventual consistency. With the concept of consistent adaptations, we solve the problem of an inconsistent configuration during the execution of adaptations on the actors. No central actor is needed to supervise the adaptations, and the nodes coordinate to bring the system from the source role configuration to the target configuration. As already discussed in Sect. 1, this atomic transactional approach comes with downsides regarding the availability of actors in case of a slow adaptation process or failure. Therefore, we present a concept for an approach for eventual consistent adaptation execution so that rollbacks are not necessary, and nodes that failed can be recovered after a short time. Instead of reverting all states as they were before the adaptation process started, intermediate changes can already be run, and some parts of the system can already profit from potential improvements.

## *4.1 Rolactor DSL*

The domain captured by our proposed DSL is a system composed of actors which can dynamically start and stop playing roles in structured contexts adhering to the CROM meta-model. In the following, we discuss the models that can be defined using the language - I.e., what parts a Rolactor application consists of. The structural meta-model for the language comprises the following top-level elements:

– A *natural type* is an active object type that corresponds to the concept of a natural object with inherent identity in the CROM meta-model. Thus, natural implies that an object of this type can be created and exist outside of any context. Natural objects can define a set of message handlers and internal state and a set of natural object types that can be nested inside an object of this type. Message handlers contain embedded (Java) code. Internal state may consist of active object types. Also, relationship partners can be defined as bidirectional relationships or unidirectional relationships. Relationship partners can be of any active object type. Additionally, an *init* message handler that is asynchronously invoked can be added when creating an object of this type.

```
Natural = “natural” , “actor” , Identifier , “on” ,
Identifier , “{“ , Component* , “}”
Component = Handler | State | BiEndpoint | Child |
  UniEndpoint
Handler = “handler” , Type , Identifier , “(“
  Parameter* , “)” , “{“ , Statement* , “}”
State = Type , Identifier
BiEndpoint = “rel” , Identifier , “->” ,
  Identifier , ".", Identifier , "[", Cardinality,
   “]”
UniEndpoint = “ref” , Type, Identifier,
  Cardinality, “]”
Cardinality = “1” | “n” | “0..n” | “0” | “0..1”
Child = “contains” , Identifier , “[“, Cardinality,
  “]”
```

A simple active object definition in the language may look as follows:

```
natural actor HelloActor on Server {
  contains ChildActor[1]
  ref Console console;
  init(){
    // Embedded code
  }
  handler string SayHello() {
    return “Hello World”;
  }
}
```

Here, an active object type *HelloActor* is defined, which contains exactly one child actor, references an actor of type *Console,* and defines two message handlers: one init message handler called asynchronously on object creation, and a *SayHello* message handler which just returns a hello world message.

– A *node* defines a type of node that can participate in the distributed application, e. g. a client or a server. Each node type defines a natural type of which one object is created at the application startup to serve as an entry point for the application. A *Rolactor* Program may consist of any number of node types and is compiled to a single Java executable that can subsequently be deployed to participating nodes. The node type is passed as a command line parameter to the local application upon launch.

– A *role type* is an active object type corresponding to role objects in the CROM meta-model. As such, a role object cannot exist on its own since it depends on the existence of an object that plays it in a compartment. Apart from this, a role type can contain all concepts discussed for natural types - I.e., message handlers, internal state, relationship partners, etc.

```
Role = “role” , “actor” , Identifier , “playedBy” ,
  Identifier+ , “on” , Identifier , “{“ ,
  Component* , "}"
```

– A *compartment type* is an active object type containing both role and natural types. For each role type, player types can be defined.

```
Compartment = (“natural” | “role”) , “compartment” ,
  “actor”, Identifier , “{“ , CComponent* , “}”
CComponent = RoleContainment | RoleConstraint
RoleContainment = “role” , Identifier
```

Additionally, compartments can contain role constraints.

```
RoleConstraint = Identifier
  (“prohibits” | “implies”) , Identifier ,
  (“before” | “after”), (“itself” | Identifier)
```

**Message Handlers.** As said, message handlers can be invoked asynchronously by active objects in a relationship with the object. Received messages are processed sequentially in the order of arrival. Message handlers define a list of parameters, optionally a return type (of which none may be of an active object type) and contain any number of code statements that may mutate internal state or send messages to other active objects (or the object itself). We assume an embedding of Java as a base language for statements, excluding object creation. Rolactor extends Java expressions by two binary operators:

– A *fire-and-forget* operator *“!”* can be used to send a message to another active object. A fire-and-forget expression does not return a value, and execution of the message handler is continued immediately. The fire-and-forget operator works like method invocations in imperative programming languages, apart from its asynchronous execution.
– An *ask* operator *“?”* can be used to request a response from another active object. An ask expression returns a *future* value, which represents a value that will be received asynchronously in the future. To access the future value, the language supports an await expression, which suspends the current message handler’s execution, saves the message handler’s execution state and local variables, and

allows the active object to process other messages. When an answer that completes the future sent is received, the execution of the message handler is resumed.

As roles in Rolactor adapt active objects, they can introduce new behavior (i.e., message handlers) or relationships to their players. Still, they may also alter the player's behavior by intercepting messages. For this purpose, message handlers in roles can be annotated with an *intercept* clause, which defines a message handler in the player. If an active object then sends a message to the player object while the role is played, the intercepting message handler is executed instead of the player's handler. The role can forward the message to the intercepted handler by using the keyword *base* as a parameter in *fire-and-forget* or *ask* expressions. Thus, an *ask* expression combined with *await* can be used to alter the response to an intercepted message.

**Constraints.** As mentioned, compartments can also contain binary role constraint definitions. The following types of constraints are supported:

- An *implies* definition conveys that a role implies another role. If a player plays a role of type *A* in a compartment *C,* which contains a role constraint that defines that *A* implies a role type *B,* the same player object must also play a role of type *B.*
- A *prohibits* constraint is the opposite of an *implies* constraint, so a player playing *A* may not play *B* at the same time.
- *Equals* and *unequals constraints* are shorthand for two constraints "*A implies B"* and "*B implies A",* and "*A prohibits B"* and "*B prohibits A"* respectively.
- Implies and prohibits can be limited to a positioning definition. This is only relevant if *A* and *B* both intercept messages of the same type – in this case, *A* can prohibit role *B* from being played before itself to intercept messages first. If *A* then forwards the message using the *base* keyword, it will be intercepted by *B* before the player can process it.
- An *after* constraint works the other way round.

We can extend the previous hello world example with a role that adds some text to the message returned by the *SayHello* handler:

```
role actor NiceHelloActor playedBy HelloActor on Server {
  handler string Hello() intercepts Hel-
loActor.SayHello(){
    return (await base ? SayHello())
      + "and have a good day";
  }
}
natural actor NiceCompartment on Server {
  role NiceHelloActor;
}
```

This example combines the *await* expression with the *ask* and *base* expressions allowed by the language; the intercepted *Hello* message is asynchronously forwarded to the player, and the subsequent result is awaited. Afterward, another message is appended, and the result is returned.

If we have multiple roles played by *HelloActor*, we can use constraints to control which role should intercept a message first:

```
role actor LoudActor playedBy HelloActor on Server {
  handler string Hello() intercepts HelloActor.SayHello(){
    return (await base ? SayHello()).toUppercase();
  }
}
natural compartment actor ExampleCompartment on Server {
  role NiceHelloActor, LoudActor;
  LoudActor prohibits NiceHelloActor before itself;
}
```

Here, we have defined that *HelloActor* can play both roles *NiceHelloActor* and *LoudActor* in the *ExampleCompartment* context at the same time. Since the *LoudActor* transforms the message to uppercase, it should intercept the *Hello* message to the base actor after the other role so that the complete message is transformed. This is expressed by a *prohibit* constraint inside the compartment, which defines that as long as an object plays the *LoudActor* role, it may not play the *NiceHelloActor* role before the *LoudActor* role so that *LoudActor* would intercept the message first.

In the next section, we extend the DSL with a transaction concept that enables the consistent adaptation of the system. By using transactions, new active objects can be created or removed, and relationships between objects can be established.

## *4.2 Consistent Adaptation with Adaptation Transactions*

Weißbach et al. introduced a mechanism for coordinated adaptations in decentralized self-adaptive systems. Their approach follows a strict transactional method, where all changes in an adaptation transaction get executed, or intermediate changes get reverted in case of an error. As a first step, we want to apply that methodology to our approach and integrate respective concepts into the DSL.

The architecture of the execution component looks as depicted in Fig. 5.

Every actor contains a so-called adaptation manager (AM). This component coordinates the current adaptation transaction and communicates with all other AM on the other actors. An adaptation transaction is a unit that contains multiple operations, so-called adaptation operations. An adaptation operation contains information on which change is executed on a node, e.g., adding or removing a role from a player.

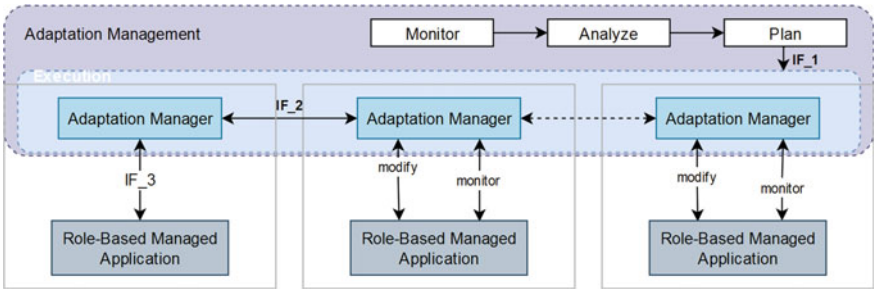



**Fig. 5** Adaptation of the encryption during run-time [10]

All adaptation operations in an adaptation transaction succeed, or all intermediate operations will be rolled back to keep a consistent state. In our DSL, transactions define the active object type that may initiate them, a list of parameters (which may not be of active object type), and a list of internal variables that can be used to reference newly created active objects. They may contain any number of the following operations:

- *Create* and *remove* operations can be used to create new natural active objects or remove them. The parent object and the node on which the active object should be run have to be defined for each *create* operation.
- *Play* and *drop* operations are used to *create* new role active objects or remove them. The player object, the target compartment, and the node on which the object should be run have to be defined for each *play* operation.
- *Connect* and *disconnect* are binary operations that establish relationships between objects. The two relationship endpoints defined in the objects have to be set for each operation.

A transaction can be written as follows:

```
  Transaction = "transaction" , Identifier, "(" , Parame-
ter+
    , ")" , "initiatedBy" , Identifier , "{" Variable+
    Operation+ , "}"
  Parameter = Type , Identifier
  Variable = Type , Identifier
  Operation = NaturalOperation | RoleOperation |
    ConnectOperation
  NaturalOperation = Identifier ("create" | "remove")
Identifier "on" Expression
  RoleOperation = Identifier ("play" | "drop") Identifier
"on" Identifier "in" Identifier
  ConnectOperation = Identifier ("connect" | "disconnect")
Identifier
```

A simple transaction that creates a new natural compartment active object and a role object played within may look like this:

```
transaction CreateRole(Actor a, Actor parent) initiatedBy
  anotherActor {
  NiceHelloActor role = new NiceHelloActor();
  ExampleCompartment compartment = new
    ExampleCompartment();

  parent create compartment on parent.node;
  a play role in compartment on parent.node;
}
```

Transactions are initiated asynchronously by *adapt* statements inside of message handlers in active objects. Before the adaptation operations defined inside the transactions are executed, we must bring the affected application parts into a safe state to avoid message loss, data loss, and invalid application states and errors. To achieve this, we rely on the concept of quiescence which was first introduced by Magee et al. [7]. It is a disruptive technique since a quiescent node is not allowed to perform any task during the passive state. Additionally, no future tasks can be planned for the node while it is passivated, and all current computations must be finished before entering the quiescent state. In our approach, we define that an active object is quiescent if the following criteria are met:

- The object is not processing any messages and will not do so in the future.
- No objects the object is in a relationship with, and no role objects played by the object are processing any messages or may do so in the future.

To determine whether an object will process messages in the future, the *future* abstraction of our DSL is used. If an object is still waiting on the completion of a *future*, it cannot be quiescent.

Once quiescent, all changes can be performed, including removing or adding nodes into the application. All participating nodes are activated and start working again if all adaptation operations succeed. In case of an error all intermediate changes are rolled back, and all passivated units are activated in that case, too.

## *4.3 Eventual Consistency for Self-Adaptation*

Although we preserved a consistent system configuration with the concept presented in Sect. 4.2, it has several drawbacks in various use-cases where temporary inconsistencies can be tolerated. The atomic approach requires all operations to succeed. Otherwise, we must roll back every intermediate change, which is time and resource-consuming. If already adapted entities could rerun their tasks regardless of the outcome at other nodes and actors, the global downtime of the system can be reduced.

That approach does not apply to every application but applies to enabling compression in our use case.

Cases which could cause an adaptation error are manifold. The self-adaptation execution can be distributed due to network or connection errors. Furthermore, bringing the application to a quiescent state can also take time for some actors. For this reason, we consider already executing changes for some actors before all devices are ready. This accelerates the adaptation process so that a subset of nodes and actors can already profit from those changes instead of waiting.

For this purpose, we extend the transaction concept in our DSL to allow certain operations to be marked as *relaxed*. Relaxed operations may be executed even if other operations fail.

Our approach is inspired by a concept for implementing eventual consistency known from the Dynamo Key-Value Store [22]. To ensure that every node in a distributed database gets missing information about database updates in case of a disconnection or other errors, they rely on a so-called **"hinted handoff."** The InfluxDB[1] also adopts that concept. Simplified, their approach works as follows. If a node is not reachable or does not update the records contained in a transaction, the information is stored on behalf of the erroneous node by the other nodes in the system. The so-called "hint" is stored in a separate database. When the failed node is detected to be active again, the node gets the hint that an update is pending. The stored write operations are applied immediately to recover consistency.

Dynamo does not rely on a traditional quorum approach [22]. DeCandia et al. use a so-called "sloppy quorum" to deal with server failures and network partitions" [22]. Write and read operations on their nodes can be described using a tuple with three elements (N, R, W), which the systems engineer can set. N denotes the number of replicating nodes, R the number of successful reads, and W the number of successful writes to determine if a get or put operation was successful. To deal with partitioning and failures, low values for the W (1 for the highest availability) parameter can be set. That means that it can be sufficient for only one write out of n potential writes to succeed and not to abort the whole transaction. The n-1 remaining in case of a failure can be recovered by relying on the successful write to the last node alive. The write requests get only rejected if every node is unavailable for a write operation. For our actor-based system, we adopt the approach in a decentralized way and modify that concept for adaptation execution.

We developed a hinted handoff approach for adaptation operations and adaptation transactions. Every actor in our system has storage for incomplete operations of other actors. The adaptation execution for eventual consistent operations works similar to the atomic approach presented in Sect. 4.2. The algorithm differs when an error for an adaptation bundle occurs. Instead of rolling back all intermediate changes of the successful operations, they can be activated. The following messages are exchanged to recover from an inconsistency:

[1] https://docs.influxdata.com/enterprise_influxdb/v1.9/concepts/clustering/ (10.06.2022).

- *SendHint(operation):* As soon as the initiator recognizes that another actor failed, the execution of an adaptation or disconnected and did not receive the adaptation message, the initiator sends a hint to all other available actors.
- *SendAliveMessage():* A reconnected actor loads its durably stored network configuration and sends an alive message to one of the peers, which then initiates the compensation process
- *IncompleteRecovered(operationID):* When an operation was successfully executed and the failure is compensated, the hint concerning a specific operation can be deleted at all storage locations

*Clientactor1* detected a weak connection and wants both servers to adapt and change their role to "compressed," which is depicted in the circled "1" in Fig. 6. *Serveractor2* successfully finished the adaptation, whereas *serveractor1* failed because it disconnected for a short time because of network congestions. The initiator of this adaptation is now responsible for sending the information that an incomplete adaptation operation exists to all other actors in the system by sending a *sendHint(operation)* message, which is number 2 in Fig. 6. All actors who received the message, including the initiator, store the pending operation locally as a hint in a local storage. In our example, *serveractor1* recovers from its disconnected state and loads the internal network configuration which is stored durably. Then, an *aliveMessage()* is sent to one of the peers, in our example *clientactor1*, to get information about possibly missed adaptations. The contacted actor checks the hint storage if a hint for the node is present. If so, our contacted *clientactor1* loads the hint from the local storage and sends a *sendHint()* message containing the pending operation to the reconnected peer. *Serveractor1* executes the pending change, restores the global consistency, and notifies the other actors about the success of the compensation task. When the other peers receive the success message, they can safely delete the hint from their local storage.

It is possible that during the compensation phase, in which hints for incomplete operations exist, new adaptation operations occur for actors who must still be adapted by older operations. Therefore, the default case in our concept demands that compensation must be executed first before later adaptations can be executed because operations can depend on previous operations. Thus, the local operation order must be preserved to ensure the correct results of the adaptation execution. The MAPE-K feedback loop planning component might issue adaptation plans meant to be executed on the synchronized, i.e., recovered state.

## 5 Empiric Case Study

In this section, we want to apply the concepts presented in Sect. 4 to the distributed file server case study. We divide the case study evaluation into two parts, to first show the case with a strict atomic approach, and second, to show the applicability of the eventual consistency paradigm on self-adaptive software systems.

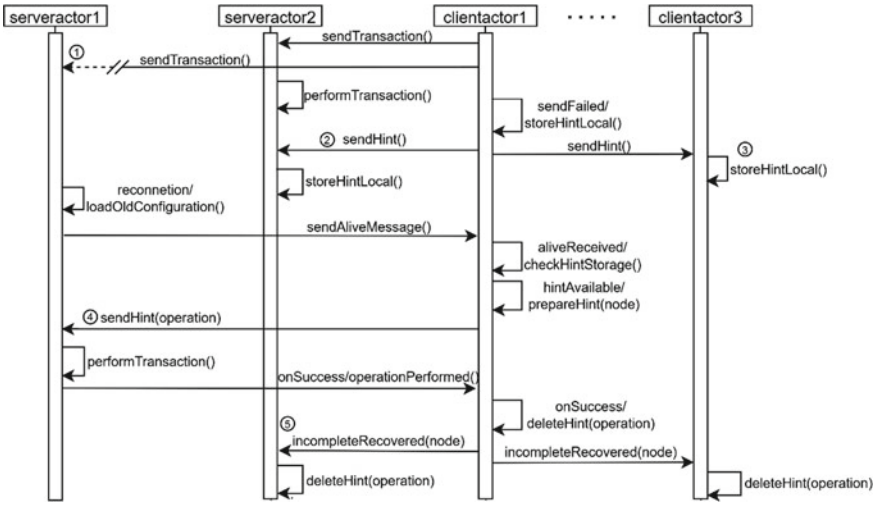



**Fig. 6** Sequence diagram for eventual consistency

## *5.1 Transactional Approach for Self-Adaptation Execution*

In our example, the participating server and client have been defined as nodes in the Rolactor DSL. On the participating node, a file server actor manages a directory of files servable to clients. After a connection between server and client has been established, a transaction is used to establish a relationship between the file server actor and an actor on the client node that wants to access a file. At some point afterward, a managing actor on the server decides that the connection should be encrypted and initiates a role play on both the file server actor and the client actor:

```
transaction EnableEncryption(fs FileServer, c FileClient,
  SecurityContext compartment, String encryptionKey) {
  Encryptor enc = new Encryptor(encryptionKey);
  Decryptor dec = new Decryptor(encryptionKey);

  fs play enc on fs.node in compartment;
  c play dec on c.node in compartment;
}
```

Here, *FileServer, FileClient,* and *SecurityContext* are active object types, of which some instances are passed as parameters to the transaction. The latter represents a compartment in which the former two play the encryption and decryption roles, which are created as active objects inside of the transaction. The encryptor could be implemented as follows:

```
role actor Encryptor on Server {
  String key;
  init(String key){
    this.key = key;
  }
  handler byte[] read(int offset, int length) intercept
    FileServer.read {
    return aesEncrypt(await base ? read(offset, length),
      key);
  }
}
```

The role uses the message interception mechanism discussed in Sect. 4.1. After the FileServer actor starts to play the role, all calls to the read handler that the client calls to retrieve the file are intercepted by the role, which forwards the message to its base and afterward encrypts the content. The ask operator "?" is used in combination with the *await* expression to forward the message to the base object asynchronously.

## 5.2 *Eventual Consistent Approach for Self-Adaptation Execution*

Recap our case study from Sect. 2. We introduced image compression for our file server example, which can be requested by every node in the cluster. In contrast to the encryption example, where it is crucial that every participant activates or turns off encryption to ensure secure communication and that received packages are interpreted correctly, receiving uncompressed images instead of compressed ones is not a crucial failure and can be tolerated in our application. The major drawback is a slower file transfer and a possibly impaired user experience for the recipient, but the system's integrity is still preserved and does not lead to a crash.

The scenario of image compression enables the usage of the eventual consistency concept for adaptation operations which was presented in Sect. 4.3. In contrast to a transactional approach, we can define operations to be executed relaxed so that no rollback is needed in case of an error. Instead, intermediate changes can get activated and start running, so already adapted nodes profit from the image compression before all entities execute the change. The remaining nodes must be recovered in subsequent steps to bring the whole system to a consistent state again.

Like in the example in 5.1, a client issues an adaptation plan to react to new circumstances because of a changed context. The monitoring sensors of a client detect a weak connection of an actor and cannot receive large files in a reasonable time to satisfy the user experience. According to the MAPE-K feedback loop, the monitored data is analyzed by the analysis component, and the planning stage generates a plan for how to overcome this situation. The planner reasoned that the server should

send compressed images to receive images faster. Therefore, a planner invokes a transaction that could be written as follows in our *Rolactor* DSL:

```
transaction EnableCompression(s1 Server, s2 Server,
  comp Compartment) {
  CompressingActor c1 = new CompressingActor();
  CompressingActor c2 = new CompressingActor();

  s1 play c1 on s1.node in comp relaxed;
  s2 play c2 on s2.node in comp relaxed;
}
```

The listing shows a transaction that can be executed in an eventually consistent way. If executed together, the two actors who work as file servers for the client actors can ensure the best availability of their services. On success at both servers, every client response results in the desired way with a compressed image. If one fails, clients requesting images of the non-adapted server will receive original-sized images. Nevertheless, in this setting, at least one updated server is more beneficial than none from a global point of view. Using our compensation approach presented in Sect. 4.3, we can deal with that temporal inconsistency by storing the pending operations on behalf of other nodes, so those operations can be executed as soon as the crashed node reconnects.

## *5.3 Limitations of the Concept*

In Sect. 4.1, we presented our concept for a DSL with role-playing actors to solve the concurrency problem and distribution for role-based systems. In addition, our DSL supports adaptability, which is beneficial in dynamic environments. We complemented this DSL with techniques for the consistent adaptation of actors and developed a concept to increase the availability of actors during the adaptation phase in unstable environments. Although our novel DSL facilitates the development of SAS with roles, certain aspects remain unaddressed, which will be discussed in this section.

In our current version, some of the role features identified by Kühn et al. are not supported by our DSL. This includes role groups and constraints on relationships between roles. Regarding the distribution aspect, the DSL captures what objects can reside on which nodes in the system and allows to specify where to place them when created in transactions. Still, the discovery of other nodes is not customizable. To address this, we aim to extend the meta-model for node types in future language versions. Additionally, we aim to support the implicit triggering of transactions by integrating rewriting rules on the role graph [23].

Also, the approach for relaxed consistency of adaptation execution has open issues and challenges which need to be tackled. The largest downside we consider

is a missing guarantee that an eventual consistent compensation is guaranteed to be compensated in a reasonable time. In our concept, we wait until a node reconnects or prepares for compensation to recover a consistent global state. Nevertheless, it could be possible that a node will be offline for a long time, or the compensation itself fails. For that case, a timeout-based emergency rollback must be implemented to recover to an older state. To accomplish that, a distributed checkpointing mechanism or detailed logging could be used to revert all intermediate operations.

Furthermore, complex adaptation operations such as the migration or cloning of roles require a more sophisticated compensation algorithm and are not supported by the current concept for eventual consistency. The algorithm only supports primitive operations, such as adding or removing roles from a player. To solve this, we aim to improve our compensation mechanism through improved coordination.

## 6 Related Work

In the following section, we line out approaches established in the literature that accomplish a similar contribution compared to our solution. First, we give an overview of existing frameworks and strategies for decentralized self-adaptation. Here, we mainly focus on self-adaptation for role-based systems since our concept relies on that paradigm. Secondly, we discuss approaches for the consistency of adaptations. To the best of our knowledge, the research body of SAS lacks sophisticated research regarding adaptation execution in unstable environments. Therefore, we discuss prominent examples from the database domain regarding eventual consistency, which we consider as interesting for solving the problem of an eventual consistent adaptation execution in SAS. Last, we consider existing role-based programming DSLs related to our own DSL presented in Sect. 4.1.

The different categories and application areas of self-adaptive software systems are manifold. Therefore, we focus on the execution of adaptation in this section, especially on decentralized and role-based approaches. Weyns et al. [3] investigated different patterns in the distribution of the MAPE components in decentralized systems and discussed their advantages and drawbacks. Although we can apply those architectural patterns to our system, no concrete implementation hints are discussed in their work. Nevertheless, their work contains important aspects of different MAPE-K patterns' overall advantages and disadvantages. The HELENA [8] approach combines self-adaptive software systems with the role paradigm. They present a holistic engineering approach for SAS by separating adaptation and application logic. Their key concept is roles that can be adopted by components dynamically. HELENA focuses on the modeling aspect of SAS, which results in missing solutions about the run-time aspects of their approach regarding the consistency of adaptation execution. Taing et al. [1] presented an approach for consistent unanticipated adaptation at run-time in role-based applications, allowing asynchronously replacing adaptation entities. They rely on a transactional approach to ensure consistency during execution, inspired by the quiescence [7] and tranquility [6] concepts. Their

dynamic instance binding mechanism makes it possible to exchange entities during run-time. Nevertheless, there are no assumptions about distributed and decentralized adaptations since their approach aims for consistent adaptation to a local runtime. The SASSY [24] framework is a model-driven framework for self-adaptation which supports the decentralization of adaptation. They use software adaptation patterns to describe steps to adapt the system from one configuration to another. Nonetheless, they do not consider unstable channels and failures throughout the system's run-time, i.e., they assume their framework to be run in a perfect environment, which does not meet our requirements of possible network congestions. DECIDE [25] is an approach to decentralizing control loops in mission-critical applications. They perform a quantitative verification at runtime to check if components contribute to the system QoS requirements in the presence of component failures and changes. Although they consider that single nodes can fail during runtime, their focus lies on the "how" and "what" to adapt to fulfill the overall system goal if a node fails. On the other hand, we focus on the coordination and distribution of decentralized adaptations and consistent execution of changes, i.e., we take failures of the adaptation process itself into consideration, which is not handled by their approach. The approach from Weißbach et al. [10] is the closest related work to our concept. Their concept combines decentralized self-adaptation with a role-based approach as the underlying managed application. They introduced a method to coordinate adaptations among decentralized peers with a strict transactional approach. Coordination messages between peers participating in the current adaptation transaction are exchanged to safely execute the transaction, like in the two-phase commit protocol (2PC). If one operation in the transaction fails, all intermediate changes are rolled back to the initial state. Due to the strict atomicity, their approach is not applicable for adaptations where temporal inconsistencies can be tolerated. Moreover, the transactional approach decreases the availability of the system during the whole adaptation process if it comes to failures or delays.

One of the aspects of our work is maintaining the system's consistency after adaptations were performed. As mentioned, Weißbach et al. introduced an approach similar to 2PC for adaptation transactions. On the one hand, consistency is preserved at every point. On the other hand, use cases exist where parts of the system can already run the updated version, and compensation tasks can be executed to recover consistency later. Tomforde et al. [9] argue that the CAP-theorem, known from the distributed systems domain, also applies to SAS. 2PC and the coordination protocol from Weißbach et al. maintain consistency and availability of the system. Still, adaptations cannot be performed in the case of system partitioning due to errors or network congestions. To support the system availability and network partitioning, consistency must be sacrificed. Current approaches such as Conflict-Free Replicated Datatypes (CRDT) [11] are a solution to increase availability and allow changes to be performed in a partitioned environment by sacrificing consistency for a short time. CRDTs apply the paradigm of eventual consistency, known from the database domain. They can theoretically be adopted as a technique for model synchronization for adaptive systems. However, we aim for a step-by-step recovery of consistency in the case of compensation tasks to a specific global state. CRDTs converge to an overall global consistent state. However, this final state could be a state we did not aim for since

synchronization in CRDT is achieved in a commutative way, and the CRDT's nature does not foresee influencing the merge process manually.

A number of domain-specific languages for role- and actor-based programming have been proposed in the literature. Notable examples of role-oriented programming include Object/Teams Java [16], SCROLL [17] and the LyRT DSL [18]. While Object/Teams follows the notion of teams that group roles together, SCROLL and LyRT adhere to the concept of reified structured context and compartments. However, all three DSLs do not consider the aspects of concurrency and distribution. The closest related work to our Rolactor programming language is the ContextErlang approach [23], which enables context-oriented programming with actors in Erlang. Compared to our work, ContextErlang does not feature roles and does not consider coordination of adaptations using a transaction concept or other means. Recent examples of DSLs that support programming with active objects are Encore [19] and SALSA [20]. However, both do not consider context-specific adaptations of objects and roles.

## 7 Conclusion and Future Work

In this work, we presented a role-based framework for concurrent adaptation execution using the actor model of computation. We discussed the limitations of current role frameworks, such as LyRT [18], Scroll [17], and OT/J [16]. We presented a way to overcome the limitations of implementing distributed role-based applications with the existing frameworks. We demonstrated a concept to execute adaptations consistently to keep a globally consistent view of the configuration and identified the limitations of a strict transactional approach. Some applications do not require a transactional approach for adaptation operations and can continue working even if single adaptations fail. Therefore, we introduced a concept for an eventual consistency approach in SAS for that case. Our approach was evaluated empirically using a distributed file server application case study that can send encrypted or decrypted data and compresses data if requested by the client actors.

In the future, we aim to do performance benchmarks for the system to compare our approach against current state-of-the-art role runtimes. Furthermore, we want to further develop and model-check the eventual consistency algorithm for adaptation execution to show the deadlock freeness of this approach.

## References

1. Taing N, Wutzler M, Springer T, Cardozo N, Schill A (2016) Consistent unanticipated adaptation for context-dependent applications. In: Proceedings of the 8th international workshop on context-oriented programming, COP 2016, pp 33–38. https://doi.org/10.1145/2951965.2951966
2. Kephart J, Chess DM (2003) The vision of autonomic computing. Computer 36:41–50. https://doi.org/10.1109/MC.2003.1160055

3. Weyns D, Schmerl B, Grassi V, Malek S, Mirandola R, Prehofer C, Wuttke J, Andersson J, Giese H, Göschka KM (2013) On patterns for decentralized control in self-adaptive systems. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 7475 LNCS, pp 76–107. https://doi.org/10.1007/978-3-642-35813-5_4
4. Shmelkin I, Matusek D, Kluge T, Springer T, Schill A (2022) Intent-based adaptation coordination of highly decentralized networked self-adaptive systems. In: Klymash M, Beshley M, Luntovskyy A (eds) Future intent-based networking. Lecture notes in electrical engineering, vol 831. Springer, Cham. https://doi.org/10.1007/978-3-030-92435-5_5
5. Cheng SW, Garlan D, Schmerl B (2005) Making self-adaptation an engineering reality. In: Babaoglu O, Jelasity M, Montresor A, Fetzer C, Leonardi S, van Moorsel A, van Steen M (eds) Self-star properties in complex information systems. Springer, Heidelberg, pp 158–173
6. Vandewoude Y, Ebraert P, Berbers Y, D'Hondt T (2007) Tranquility: A low disruptive alternative to quiescence for ensuring safe dynamic updates. IEEE Trans Softw Eng 33:856–868
7. Kramer J, Magee J (1990) The evolving philosophers problem: dynamic change management. IEEE Trans Softw Eng 16:1293–1306. https://doi.org/10.1109/32.60317
8. Klarl A (2015) Engineering self-adaptive systems with the role-based architecture of Helena. In: Proceedings - 2015 IEEE 24th international conference on enabling technologies: infrastructures for collaborative enterprises, WETICE 2015, vol 2, no 4, pp 3–8. https://doi.org/10.1109/WETICE.2015.32
9. Tomforde S, Gruhl C (2020) Fairness, performance, and robustness: is there a CAP theorem for self-adaptive and self-organising systems? In: Proceedings - 2020 IEEE international conference on autonomic computing and self-organizing systems companion, ACSOS-C 2020, pp 54–59. https://doi.org/10.1109/ACSOS-C51401.2020.00029
10. Weißbach M, Chrszon P, Springer T, Schill A (2017) Decentrally coordinated execution of adaptations in distributed self-adaptive software systems. In: Proceedings - 11th IEEE international conference on self-adaptive and self-organizing systems, SASO 2017, pp 111–120. https://doi.org/10.1109/SASO.2017.20
11. Preguiça N, Baquero C, Shapiro M (2018) Conflict-free replicated data types (CRDTs). In: Sakr S, Zomaya A (eds) Encyclopedia of big data technologies. Springer, Heidelberg. https://doi.org/10.1007/978-3-319-63962-8_185-1
12. Steimann F (2000) On the representation of roles in object-oriented and conceptual modelling. Data Knowl Eng 35(1):83–106. https://doi.org/10.1016/S0169-023X(00)00023-9
13. Kühn T, Leuthäuser M, Götz S, Seidl C, Aßmann U (2014) A meta-model family for role-based modeling and programming languages. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 8706. Springer, Cham, pp 141–160. https://doi.org/10.1007/978-3-319-11245-9_8
14. Bachman CW, Daya M (1977) The role concept in data models. In: Proceedings of the third international conference on very large data bases, vol 3, pp 464–476
15. Hewitt C, Bishop P, Steiger R (1973) A universal modular ACTOR formalism for artificial intelligence. In: Proceedings of the 3rd international joint conference on Artificial intelligence (IJCAI 1973). Morgan Kaufmann Publishers Inc., San Francisco, pp 235–245
16. Herrmann S (2003) Object teams: improving modularity for crosscutting collaborations. In: Aksit M, Mezini M, Unland R (eds) Objects, components, architectures, services, and applications for a networked world. NODe 2002. Lecture notes in computer science, vol 2591. Springer, Heidelberg. https://doi.org/10.1007/3-540-36557-5_19
17. Leuthäuser M, Aßmann U (2015) Enabling view-based programming with SCROLL: Using roles and dynamic dispatch for establishing view-based programming. In: Proceedings of the 2015 joint MORSE/VAO workshop on model-driven robot software engineering and view-based software-engineering (MORSE/VAO 2015). Association for Computing Machinery, New York, pp 25–33. https://doi.org/10.1145/2802059.2802062

18. Taing N, Springer T, Cardozo N, Schill A (2016) A dynamic instance binding mechanism supporting run-time variability of role-based software systems. In: Companion proceedings of the 15th international conference on modularity (MODULARITY Companion 2016). Association for Computing Machinery, New York, pp 137–142. https://doi.org/10.1145/2892664.289268
19. Brandauer S, Castegren E, Clarke D, Fernandez-Reyes K, Johnsen EB, Pun KI, Yang AM (2015) Parallel objects for multicores: a glimpse at the parallel language encore. In: International school on formal methods for the design of computer, communication and software systems. Springer, Cham, pp 1–56
20. Varela C, Agha G (2001) Programming dynamically reconfigurable open systems with SALSA. ACM SIGPLAN Notices 36(12):20–34
21. De Koster J, Van Cutsem T, De Meuter W (2016) 43 years of actors: a taxonomy of actor models and their key properties. In: Proceedings of the 6th international workshop on programming based on actors, agents, and decentralized control (AGERE 2016). Association for Computing Machinery, New York, pp 31–40. https://doi.org/10.1145/3001886.3001890
22. De Candia G, Hastorun D, Jampani M, Kakulapati G, Lakshman A, Pilchin A, Sivasubramanian S, Vosshall P, Vogels W (2007) Dynamo: amazon's highly available key-value store. In: SOSP 2007, pp 205–220
23. Kluge T (2020) A role-based architecture for self-adaptive cyber-physical systems. In: Proceedings of the IEEE/ACM 15th international symposium on software engineering for adaptive and self-managing systems. Association for Computing Machinery, New York, pp 120–124. https://doi.org/10.1145/3387939.3391601
24. Menascé DA, Gomaa H, Malek S, Sousa JP (2011) Sassy: a framework for self-architecting service-oriented systems. IEEE Softw 28(6):78–85. https://doi.org/10.1109/MS.2011.22
25. Calinescu R, Gerasimou S, Banks A (2015) Self-adaptive software with decentralised control loops. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 9033, pp 235–251. https://doi.org/10.1007/978-3-662-46675-9_16

# Ontology-Driven Approach to Scientific and Educational Information Representation




**Larysa Globa, Rina Novogrudska, Bohdan Zadoienko, and Yu Junfeng**

**Abstract** The paper presents an approach to ontological model development for representing information accumulated by various scientific and educational organizations. The process of such organizations functioning is associated with accumulation of a large amount of information, which in turn is used in the process of assessing the quality of their functioning. It is proposed to provide structuring and systematization of such information on the basis of ontological approach, for its further analysis and processing. The general ontology of scientific and educational organizations information identifies several interrelated components, each of which provides a description of the processes occurring in the organization, information accumulated by the organization in the process of its functioning, as well as information on quality assessment of organization. The elements of all the components of general ontology are described. The results of proposed ontology development using the ontological solutions development platform TEDAOS are given.

**Keywords** Ontology · Scientific organizations · Educational organizations · Classes · Quality assessment · Information

## 1 Introduction

Today, there are a large number of public and private institutions providing services in various spheres of social life. The demand for such institutions depends on the efficiency of their functioning. It is for the organizations involved in the fields of

L. Globa (✉) · R. Novogrudska (✉) · B. Zadoienko (✉)
Institute of Telecommunication Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremogy Ave. 37, Kyiv 03056, Ukraine
e-mail: lgloba@its.kpi.ua

R. Novogrudska
e-mail: rinan@ukr.net

Y. Junfeng (✉)
Information Research Institute of Shandong Academy of Science, Shandong, China
e-mail: yujf@sdas.org

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_3

science and education that the efficiency of functioning is of particular importance, because the quality of services they provide depends on the future level of the state development and the life of its citizens. That is why there are currently some generally accepted principles for evaluating such organizations. The process of such peer review is outlined and approved in the relevant regulatory documents [1–3].

In the course of their operation, institutions accumulate a great deal of information (when it comes to scientific and educational organizations, such information includes scientific articles, reports, employee ratings, economic indicators, etc.) that can be used to evaluate them according to the national criteria. Thus, the problem of such information analyzing and processing arises in the process of organizations functioning assessing. The idea to systematize and structure such information using modern means of artificial intelligence (AI) can solve such a problem. Such AI means would not only facilitate information use at the institution level, but may also enable it to be used by the relevant authorities in the process of evaluating the quality of the institution's functioning.

Today there is a tendency to the development of artificial intelligence and their penetration into any sphere of human life. Specialized technologies based on artificial intelligence are used to develop information systems of various kinds and purposes. Artificial intelligence technologies also offer a wide range for the presentation of knowledge and data. Among the existing models of knowledge representation it is the ontological model that has acquired the greatest use. Presented study proposes to use the ontology as one of the models of knowledge representation to organize information of scientific institutions for its structuring and systematization, as well as for its further processing and use.

The paper describes the approach to ontological model development for the representation of information accumulated by various scientific institutions and organizations. Such an approach involves the design of the ontological model structure, the description of ontological model elements, and ontology direct development, as well as filling it with data using the appropriate environment for the ontological models development.

The structure of the paper is the following: Sect. 2 shows analyses of related works and backgrounds for the research. Section 3 gives the description of ontology system structure. In Sect. 4, elements of ontological model are depicted. Section 5 describes main features of TEDAOS software platform that is proposed to be used for ontology development. Section 6 presents conclusions and plans for future work.

## 2 Background and Basic Notions

In recent years, it has become necessary to create a user-oriented approach to describing the knowledge accumulated in the information systems due to the rap-id development and use of artificial intelligence technology for the development of such

systems. Artificial intelligence technologies involves different knowledge representation models, among them: production modes, semantic model, logical model, frame model and ontological model.

The frame model is based on the concept of a frame. Frame representation is an important knowledge representation formalism that allows one to establish and display the concept of inheritance. The frame technique includes: the number of frames or nodes linked by relationships. Each frame explains both the instance and the class frame. The idea of a frame was first introduced by Marvin Minsky in 1975 as a basic way of showing the range of knowledge.

A frame is a group of properties that identify the state of an object, and that object is related to other frames or objects. In fact, a frame is more than just a record, or perhaps a data structure containing data.

In artificial intelligence, a frame is known as a slot-filling knowledge representation method.

It is now known that slots that are instances are typically populated with an "object". In this case, the object may be a physical object. The object can be a property (such as a shape or color) or a location, scenario, or emotion.

Popular attributes in the presented knowledge are used based on the frame method:

1. The structure of frames seems to consist of frames arranged in a hierarchical form.
2. The main components of the frame are slots, and the fillers used for these slots, which must be specified.
3. Characteristics (fillers, filler constraints, etc.) tend to inherit super frames in order to use "subframes" from the structure according to several inheritance tricks.

A frame can be described by a set of slots. Each slot explains a specific feature or procedure from the frame. Slots are used to store values. A slot may include default values for various frames and a set of principles by which the actual slot values can be obtained.

Slots can be terminal or be frames, so they form a whole hierarchical network. The frame has an almost homogeneous structure and consists of standard units. Each such unit (slot) contains a name and has its own meaning [1].

Formally, a frame is a view data type:

$$F = < N, S1, S2, S3 >,$$

where N—the name of the object, S1—a set of slots containing facts that define the declarative semantics of the frame, S2—a set of slots that provide connections with other frames (causal, semantic, etc.), S3—a set of slots that provide transformations that define the frame's procedural semantics.

The composition of frames and slots in each specific frame model may be different, but within the framework of one system, a single representation is advisable to eliminate unnecessary complication [2]. The nesting depth of slots in a frame (the number of levels) depends on the subject area and the language that is implemented in the model.

A set of ordered frames is called a script. Frame structure is commonly used to "understand" natural languages, as well as to describe mechanical devices and how they work. Each frame can store information about a specific element, and the set of frames connected to each other will be a script representing a specific mechanical device.

The procedures that are attached to the frames provide a flexible, organized basis for calculations. Justification within a frame system typically begins with the system "recognizing" an object as an instance of a generic frame, and then applying the procedures triggered by that recognition. Such procedure calls may then retrieve more data or changes in the knowledge base that may apply to other procedure calls. When no more procedures apply, the system stops.

A frame is a data structure for representing some conceptual object [1].

In general, the frame model allows the relation of all properties of declarative and procedural knowledge.

The advantages of the frame-based knowledge representation method are described below:

1. Frame knowledge representation makes development easier by grouping related data.
2. Compared to other methods of knowledge representation described in the form of a development rule, the frame is flexible and intuitive in many areas of application.
3. The frame view is easy to understand and use for people who are not model developers.
4. It's not hard to add slots for new attributes and relationships.
5. Include default data and easily detect missing values.

The frame representation of knowledge has some disadvantages, described below:

1. It is difficult to use the frame model in the program, so the algorithm is necessary in the process of using the frame in the program.
2. Lack of computer software with low prices.
3. The inference mechanism is not easily handled in the frame system.

Network model of knowledge representation. Semantic networks are structured representations of knowledge that are used for reflection and inference. A wide variety of theories, models, methods and practical applications for creating and using meaning networks have appeared in different areas of academia and industry. Semantic networks consist of three components [3]:

- A syntax that defines the types of nodes and edges to be taken into account.
- An indication of the meaning or semantics that nodes, links, and the entire network can represent.
- Rules of inference.

Data structures representing semantic networks contain, at a minimum, nodes called concepts and edges between concepts. The concept is an abstract representation of ideas, thoughts and units of knowledge and meaning that people comprehend in their minds. Concepts can also be referents of objects. Examples of concepts are

"social network analysis", "collaboration" and "common practice". When concepts have natural language representations, the corresponding word or sequence of words is used as labels for nodes in semantic networks. If a link in semantic networks is typed, the link type explains the nature of the relationship between the connected nodes. Otherwise, links represent or establish some meaningful association between pairs of concepts. Link generation in semantic net-works is highly dependent on the type of input and the intended use of the network, as described later in this entry. In addition to these limitations, representations of semantic networks can be anywhere between the strictly formalized and the informal.

In many different approaches, the smallest structured unit of the semantic web is defined, as a rule, by a triple that describes the source or subject, the predicate, and the object or goal of the action. This basic structure is used, for example, to describe who did what to whom to represent event data and who said what to whom to represent communication data. Such triples can be further enhanced with background information such as spatial and temporal data, node and link attributes.

Semantic networks are the result of the translation or transformation of the data from which the networks were built. Translations turn the natural introduction of language into isomorphic, structured representations used as input to inference mechanisms. This approach has been seen in artificial intelligence as one attempt to understand natural language, and it has shown that it can be used for small nested data. Transformations are abstraction processes that are used to store and expose entities and relationships that are explicitly or implicitly present in the in-put. The goal with transformations is to reduce the dimensionality of the input data in order to fix the corresponding structural relationships and make the network data available for output [4].

Semantic networks are used for the individual and collective acquisition, organization, management and use of knowledge. Knowledge is acquired from se-mantic networks by performing reasoning and inference on the data of the network. Knowledge stored in and derived from meaning networks need not be factually correct or logical. This is due to two reasons: First, semantic networks were designed to represent what is meant by information. This meaning may differ from the true conditional content or the most likely interpretation of some information. Secondly, depending on the data from which the network was built, the semantic network can represent universal, culture-dependent, domain and individual knowledge.

Semantic networks have some advantages such as:

1. despite the variety of entities, they can be shown in one semantic network;
2. semantic systems provide a graphical view of the concepts under consideration from a single point of view, and for this reason they can be simple to implement and easy to understand;
3. the semantic web can be used as a normal application (application program) for communication between different areas of knowledge, for example, for information technology;
4. the semantic network makes it possible to implement a simple approach to the study of the problem space;

5. the semantic web provides an approach to creating branches of related components;
6. the semantic web affects the way people process data;
7. the semantic network is more natural than the logical representation;
8. the semantic network is characterized by greater cognitive adequacy com-pared to formalism based on logic;
9. the semantic network allows you to use an efficient inference algorithm (graphical algorithm);
10. the semantic web is more expressive than logic. The Semantic Web also has a number of disadvantages that often cause problems.

However, there are some disadvantages:

1. There is no difference between individuals and classes. The system is limited by the knowledge of the user with links in the semantic web. Links between nodes don't always look like functions. It is necessary to distinguish between links containing a number of links, and links that have a structural character. The same links can be used to connect three nodes to show the structure of the network.
2. There is no difference between the features that belong to a class and the features of a particular instance of the class, together with the features of the class.
3. Conventional semantics don't really exist; therefore, there is no agreed upon idea of what the proposed image design indicates. Semantic systems are usually based on techniques that change. An alternative to this problem can be both the use of conceptual diagrams and the KL-ONE knowledge representation formalism, which makes it possible to conquer semantic indistinctness in the semantic system. KL-ONE is a popular knowledge representation system in the semantic web and frame.

The logical (predicate) model of knowledge representation is based on the algebra of propositions and predicates, on the system of axioms of this algebra and its rules of inference. From predicate models, the model of first-order predicates based on terms and predicates has received the greatest distribution. All knowledge about the subject area in such a model is described in the form of logic algebra formulas or inference rules [5]. Description in the form of formulas makes it possible to present declarative knowledge, and inference rules—procedural knowledge. A domain model using a logical model can be defined in a simplified way as <domain model> = <conceptual knowledge> + <constructive knowledge> . Logical models are convenient for representing logical relationships between facts, they are formalized for their use, the model is a convenient and adequate toolkit, for example, a logic programming language [6].

Models of this type are based on a formal system defined by a quadruple of the form:

$$M = < T,\ P,\ A,\ B >,$$

where set T—is the set of basic elements of different nature, set P—is a set of syntactic rules, set A—is a set of axioms, set B—is the set of inference rules.

Natural languages can represent a wider range of knowledge, but logic allows a well-formulated subset to be expressed in computational form.

There are many types of logic; each of them is especially suited to its target application areas and has distinct computational capabilities and requirements. Logical systems vary in six dimensions from what can be considered basic logic, first-order logic.

Advantages of the logical view:

1. The logical view allows you to perform logical reasoning.
2. The logical representation is the basis for programming languages.

Disadvantages of the logical view:

1. Logical views have some limitations and are difficult to work with.
2. The logical presentation technique may not be very natural, and the inference may not be as effective.

The production model is a model based on rules that allows you to represent knowledge in the form of sentences like "If (condition), then (action)" [7]. Under the "condition" (antecedent) is understood a certain assumption—a pattern by which the search is carried out in the knowledge base, and under the "action" (consequent)—the actions performed upon the successful outcome of the search (they can be intermediate, acting further as conditions, and terminal or target, completing the operation of the system). Most often, the conclusion on such a knowledge base is direct (from data to the search for a goal) or reverse (from a goal to confirm it—to data). Data is the output facts stored in the fact base, on the basis of which the inference engine or rule interpreter is launched, iterates over the rules according to the production knowledge base (KB) [8].

The simplicity and clarity of this method led to its use in many systems. Knowledge processing systems that use knowledge representations by production rules are called productive production systems. The production system includes a database (DB), a rule base and a rules interpreter.

The rule base is an area of memory that contains a knowledge base—a set of knowledge presented in the form of rules of the form "if … then"; the database is a region of memory containing the actual data (Facts) that describe the input data being entered and the states of the system.

Databases in different systems have a different form, however, they can all be described as a group of data containing a data name, attributes, and attribute values. An interpreter is an inference engine, it is a system component that generates conclusions using a rule base and a database.

Consider the conclusions based on production rules.

The mechanism implemented today as a means of inference in a production system is not complex. It has the functions of searching in the knowledge base, sequentially performing operations on knowledge and obtaining conclusions.

Moreover, there are two ways to draw such conclusions: direct conclusions and reverse conclusions.

In direct inference, progress is made towards the goal by applying the rules sequentially to the data that is taken as the starting point. In direct inferences, one of the data elements contained in the database is selected, and if, when com-pared, this element is combined with sending the rule, then the corresponding conclusion is derived from the rule and placed in the database, or an action is per-formed that is determined by the rule, and the content of the database changes accordingly.

Most often, such inferences are called data-driven inferences, or "upward" inferences, when new results are sequentially displayed, starting from already known data.

Conclusions, in which the process moves in the direction from the set goal to the starting point, are reversible. They are also called such "top-down" or goal-oriented inferences.

The process of top-down conclusions starts from the set goal. If this goal is consistent with the conclusion of the rule, then the premise of the rule is taken as a sub-goal or hypothesis. This process continues until the subgoal or hypothesis matches the data.

It is impossible to categorically answer the question which of these methods is better, since it depends on the problem for which they are used. In systems that require high universality, it is necessary to have both methods of output.

Productions are the most popular means of representing knowledge in information systems. Any production model rule consists of one or more attribute-value pairs. The working memory of the production system stores attribute-value pairs, the truth of which is established in the process of performing a specific task. The content of working memory changes during the execution of the task [9]. When describing the knowledge of a particular subject area for resource structuring, a separate entity of the subject area is considered as an object, and the data stored in the working memory representing values take on the attributes of this object. One of the benefits of this knowledge representation is the clarification of the context for which the rules apply. Rules from the rule base can fire more than once during the same inference, since one rule can be applied to different object instances (but not more than once to each instance).

The production model is most often used for production subject areas. It is characterized by visibility, high modularity, ease of additions and changes, and simplicity of the inference mechanism.

Ontological modeling is one of the approaches of AI to identify the subject domain, based on the idea of the conceptual modeling. At present, there is no single definition for the concept of ontology. The very concept of ontology comes from the Greek "Ontos" is a being, "logos" is a doctrine, a concept, that is, it is a branch of philosophy that studies being.

One of the definitions of ontology says that ontology is an attempt at a comprehensive and detailed formalization of some area of knowledge through a conceptual scheme. Under the conceptual scheme means a set of concepts + information about the concept (properties, relations, constraints, axioms and statements about the concepts needed to describe the processes of solving problems in the selected subject area).

Among the specialists dealing with the problems of computational linguistics, the most established (classical) is the definition of ontology given by Hubert: "ontology is a specification of conceptualization" [10].

There are also a number of extended definitions of Hubert, among which are the following:

- ontology—is an explicit specification of conceptualization, where the conceptualization is a description of many objects of the subject area and the relation-ships between them;
- ontology—is knowledge formally represented on the basis of conceptualization.

Formally, the ontology consists of terms organized in taxonomy, their definitions and attributes, as well as related axioms and derivation rules:

- ontology—a formal specification divided by conceptualization, which takes place in some context of the subject area;
- ontology—a knowledge base that describes the facts that are always assumed to be true within a particular society on the basis of the generally accepted meaning of the dictionary used [11].

The classic formalized description of the ontology is represented by three:

$$O = < X,\ R,\ F >,$$

where X—a finite set of concepts of the subject area, R—a finite set of relationships between concepts, F—finite set of interpretation functions.

However, not all ontological resources available today fall under the above definition. Today, the evolution of applied information systems is moving towards increasing their intelligence. This has a significant impact on the direction of scientific and technological research related to the use of computers, and also provides society with practically significant results. However, at a certain stage of development, further development of technology by available means becomes impossible. In such periods a qualitative leap of the means used in development is required. One of such leaps in the field of AI, aimed at further intellectualization of systems of interaction with the user, was the emergence of ontologies.

Since ontologies were the response of science to the needs of his time, their appearance occurred in several areas of knowledge. Accordingly, in each of them the resources of the ontological type were formed according to their own, specific to the field of knowledge, rules.

In the design of ontologies can be divided into two areas, for some time developed separately [12].

The first direction is related to the representation of ontology as a formal system based on mathematically accurate axioms (these are resources of the ontological type, created in various fields of mathematics).

The second direction was developed within the framework of computational linguistics and cognitive science. Here ontology was understood as a system of abstract concepts that exist only in the human mind, which can be expressed in

natural language (or by some other system of symbols). This usually does not make assumptions about the accuracy or consistency of such a system.

Thus, there are two alternative approaches to creating and researching ontologies. The first (formal)—based on logic (first-order predicates, descriptive, modal, etc.). The second (linguistic) is based on the study of natural language (in particular, semantics) and the construction of ontologies on large text arrays, the so-called corpora.

Ontologies indicate a set of elements and relationships between elements that are possible or permissible in this area. For example, by creating structured representations of event data, an ontology can adopt concepts that represent who, what, when, where, why, and how an event occurs. Mapping from data points, usually words, to such conceptual classes can be predefined in dictionaries or thesauri, or they can be identified from the data. These relationships often result from:

- Syntax or grammar of basic data, such as the above three subjects, predicates and objects.
- Relevant relationships between the concepts perceived by the analyst, which, for example, are used in the methodology of sound theory.

Ontologies can also be hierarchically structured and can cause inheritance of functions from parent nodes to child nodes. An example of a hierarchical ontology with inheritance is phylogenetic trees, also known as evolutionary trees. The relationship between the elements of the ontology can be:

- Structural, for example, links from the content or term of a pointer to a piece of text or pointers between web pages.
- Logical, such as equivalence relations (“is”) and subtype relations (“is part of”). The field of artificial intelligence provides logic that can be used to model certain relationships between concepts.

Today, ontologies are used in the process of modeling and designing many information systems in different subject areas:

- E-heals [13, 14].
- E-learning [15, 16].
- E-dictionaries and thesaurus.
- Expert systems [17, 18].
- Semantic web [19, 20].
- Automatic information extraction [21, 22].
- Etc.

It is ontological models that can be used as a universal tool for integrating data and knowledge from diverse sources of information. Using ontologies allows you to structure, organize and classify information. An important characteristic of ontological moels is that its structure a priori fits into the paradigm of the semantic web, that makes it possible to automatically adjust and use the developed ontology in the Internet environment. Such a feature is in demand in terms of the development of

modern information systems, the main requirement of which is to provide online access to users in real time.

## 3 Ontology System for Scientific Institutions Information Representation

The functioning of scientific organizations is associated with certain specific features that are not characteristic of other types of institutions. For example, the important indicators of the scientific process are as follows—the number of publications, citations, various scientific indexes and ratings, participation in international projects and programs, completed scientific topics, trained specialists of different qualification levels in different specialties and specializations, and more. That is why the information produced during the operation of such institutions also reflects all levels (aspects) of such functioning. This feature must be taken into account when developing an ontological model. Also, several basic processes are involved in the process of organization of scientific activity: organization of activity of institution; definition of indicators by which it is possible to identify the level of performance of a certain type of activity in an institution; the existence of criteria for evaluating such indicators by which it is possible to evaluate the achievements of the organization and to determine whether its activity meets the requirements; organization of the institution's evaluation process itself.

The main indicators that were identified in the analysis of regulatory documentation to assess the quality of scientific and educational institutions are as follows:

1. Statistical data on the employees of the Institution:
   - Number of researchers by position.
   - Number of GDR performers and their average age.
   - Number of employees of the Institution involved in the implementation of R & D, by category of staff.
   - Number of researchers by gender and degree.
   - The average age of researchers.
   - Number of employees involved in R&D (including part-time), by level of education.
   - Membership of employees of the institution in councils.
2. General indicators:
   - National and international scientific communities.
   - Significance and relevance of the Institution for the national strategy in the field of science.
   - Regional significance and relevance of the Institution for the regional strategy in the field of science and development of the region.

- General concept and development of the Institution and information on its activities in previous years.
- Performance results of the Institution as a whole for 5 years.

3. Research:

- GDRs performed.
- List and copies of the most important publications of the Institution.
- Number of publications prepared by the staff of the Institution for 5 years.
- Scientific and educational literature.
- Electronic scientific publications.
- Problem-oriented databases.
- Popular science publications and articles.
- Publications and speeches in the media.
- Providing scientific services and solving infrastructure problems.
- Scientific consultations.
- Transfer of knowledge and technology.
- Results of creation and use of intellectual property rights for 5 years.
- List of projects of the Institution, funded on a competitive basis from national and foreign sources.
- Scientific activities and public relations.

4. International cooperation.
5. Educational activities.
6. Compliance of equipment, facilities and staffing to implement work plans.
7. Cooperation and system of scientific relations of the Institution.
8. Improving the skills of employees and career growth of young scientists.
9. Cooperation of the Institution with educational institutions.

On this basis, it is proposed that in the general ontology, several ontologies be highlighted in order to represent information accumulated by scientific institutions [23]. As a result of this selection, the overall ontology will be a system of ontologies, each describing the appropriate subprocess within both the activities of the institution itself and the process of its evaluation. The system as a whole will allow both to structure and organize the information accumulated by scientific institutions, and to organize context-independent structures for its further processing and use.

Thus ontology system is a collection of several components (see Fig. 1).

The developed system includes such ontologies as:

- Ontology of institutions activities—reflects all possible processes that occur within the scientific process within the activities of scientific institutions.
- Ontology of institutions activities organization—describes general concepts that relate to the organization of scientific activity as a whole.
- Ontology of institution activities indicators—allows to describe in detail the indicators on all aspects of the activity of the institution.

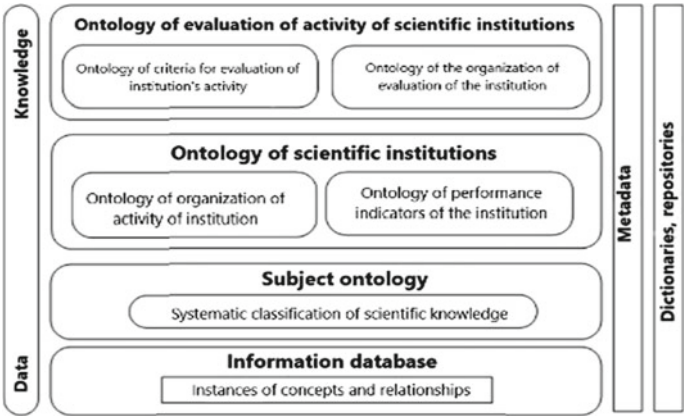



**Fig. 1** Ontology system of scientific institutions information representation

- Ontology of institutions activities evaluation—introduces concepts that make it possible to evaluate the effectiveness of the scientific activity of a particular scientific institution in the framework of its state review or establishment of its scientific level.
- Ontology of institutions activities evaluation criteria—sets criteria for evaluating the performance of the institution.
- Ontology of evaluation process organization—describes the assessment process itself by the relevant authorities.
- Ontology of subject domain—is based on the systematic classification of scientific activity.

Instances of classes and relations defined in an ontology form a database con-tent that contains terms that represent the subject domain. Initial data for the knowledge representation model that characterize the subject area are various regulatory documents, as well as textbooks, manuals, periodicals, reports, information resources and more.

Such an ontology system will not only enable the representation of information accumulated by scientific institutions in the process of activity, but also organize structures for its further use and processing.

## 4 Ontological Model Elements

In the process of constructing an ontological model, there is a need to describe its elements. The ontological model includes the following elements [24]:

$$\text{Ontological model} = < \text{classes}, \text{ attributes}, \text{ relations}, \text{ types of attribute values}, \text{ constraints on attribute values}, \text{ instances of classes} >,$$

where:

- classes—are elements of an ontological model that describe the concepts of a particular subject or problem area;
- attributes—are elements of the ontological model that describe the properties of concepts and relations;
- relations—are defined on classes, and display either the relations of classes with each other or the relations of classes to data or attributes. There are relations of the following types:
  - associative relations—allow to perform meaningful searches through the ontology information space;
  - part-to-whole relations—allow you to establish relations between classes at the hierarchy level;
  - inheritance relations—is used to pass attributes and relations from parent to daughter;
  - class-data relations—allow to associate instances of concepts with class.
- types of attribute values—specify standard types for class attribute values (for ex-ample: string, integer, real, date);
- constraints on attribute values—is used not for all attributes, but only for those whose values must lie in a certain area, they can't be less/more than a given value or they are determined by a certain rule. For example, the value of the attribute "start date" of some ontology class is constrained by T (date) = date F(T) > 0.
- instances of classes—are an ontology element that displays specific domain data that obey the structure of the ontological model.

In the course of the study, the described ontological model elements were identified for all ontologies proposed in Sect. 3 of the ontology system. The process of detailing the elements of an ontology is an important step in designing a general ontological model that will allow you to set structures for further filling the ontology with domain information (instances of classes).

## 5 Ontological Solutions Development Platform

The practical implementation of the ontological model and its filling was per-formed using the Transdisciplinary Educational Dialogues of Application Ontology Systems (TEDAOS) platform. The TEDAOS platform provides many software tools for storing and processing knowledge through the development of ontologies.

TEDAOS—provides construction of all chains of the process of transdisciplinary integration: semantic content analysis of text documents; taxonomization; highlighting the properties of taxonomy concepts; formation of the ontology of the choice problem; transdisciplinary integration of contexts based on the properties-criteria of concepts that determine the ontology of choice; inclusion of documents found in the global environment. Due to the active states of the hyper ratio of multiple partial

ordering [25], TEDAOS is an innovative IT technology for ontological management of knowledge and information resources, regardless of the standards of their creation. The functionality of IT- TEDAOS is based on information processing methods—Big Data, Data Mining, Semantic Web [26, 27]. Transdisciplinary knowledge management is based on the mechanisms of allocation of terminologies from information arrays and their taxonomization.

Thus, operationally IT- TEDAOS provides:

- network interaction with unstructured and poorly structured information of large volumes; linguistic and semantic analysis of content, aggregation and rating of information resources;
- multi-index search of thematic information with a large number of interdisciplinary links and relationships;
- adaptability to the thematic profile of each educational entity based on the Semantic Web interface with network information resources and interactive knowledge systems.

The gallery of ontological interface objects and ontograph processing (accessed via MySQL, XML) is implemented in PHP scripting language, HTML5, CSS, Java SCRIPT, jQuery visualization.

Cognitive information technology—TEDAOS, focuses on processing a large amount of heterogeneous text poorly structured information (Big Data), in automatic and automated modes, based on semantic and linguistic analysis, its structuring and classification, maintaining the process of normalized selection and forecasting, followed by analytical analysis. and network solutions. KIT-TODAOS tools provide the implementation of the following information processes:

- Linguistic and semantic analysis of network information resources, which have a significant number of relationships between disciplines, and are based on the use of various information technologies and standards.
- Transdisciplinary analysis and integration with other network information systems and WEB-oriented applications.
- Taxonomization of narratives of arbitrary documents and reflection of their structure, including connections between the context.
- Creation of ontological interactive documents.
- Detection of hidden information in the information resources being analyzed.
- Deep Learning (Machine Learning).
- Support for Semantic Web formats and protocols.
- Big Data processing.

KIT TEDAOS operational capabilities:

- Means of contextual, semantic and linguistic analysis of natural language text and construction of taxonomy of documents.
- Means of classification and generation of ontologies of the subject area.
- Means of developing a transdisciplinary ontology.

- Ontology of the choice task for information-analytical support of decision-making and provision of multicriteria analysis and rating processes.
- Means of semantic search of lexical structures based on linguistic processing of a large number of distributed network text arrays.
- Linguistic cluster—an electronic library with means of associative search of semantically connected information arrays.

Advantages. Processes of structuring and transdisciplinary integrations of distributed information resources, procedures of interaction of network information systems, which have a significant amount of interdisciplinary relations and built on the use of various information technologies and standards, and creation of expert environment for decision support are provided [28].

One of the main features of the TEDAOS software plat-form is the availability of ontological interface tools in the nomenclature. The ontological interface gives an opportunity to integrate the developed by the user ontology with network information resources and interactive knowledge systems. It provides adaptability to the thematic profile of the activity of each user subject in the TEDAOS environment. The ontological interface is implemented by the procedure of activation of multiple binary taxonomy relations. It is an intellectual means of user interaction with an ontology-based information system, that allows visualizing the results of integration and aggregation of distributed information resources in the process of organizing users communication in an easily accessible visual form [29].

An ontology for scientific institutions information representation was developed using the IT- TEDAOS environment.

In Fig. 2 shows the top-level ontology, which consists of: Rating of the institution, Rating of employees of the institution, Rating of each employee, Rating for a certain year.

In Fig. 3 shows the lower level ontology, which shows that the rating of each employee, which is divided into three areas of activity, and each area of work is divided into groups of work by type.

In Fig. 4 shows the components of each group of works, namely the individual works of the employee of the institution.

Each class of the developed ontology is described by attributes which are also set at the program level in the TEDAOS environment.

In Fig. 5 shows the attributes of the object Work No. 5. Attributes include the following data:

- Type of work.
- Points—the number of points that can be obtained for the work.
- Quantity (part made by the teacher) as a percentage.
- The result—the points received by the teacher for this work.

In Fig. 6 shows the characteristics of the work in a particular group, and general information about the work of the teacher, as well as the number of points of all works included in this group.

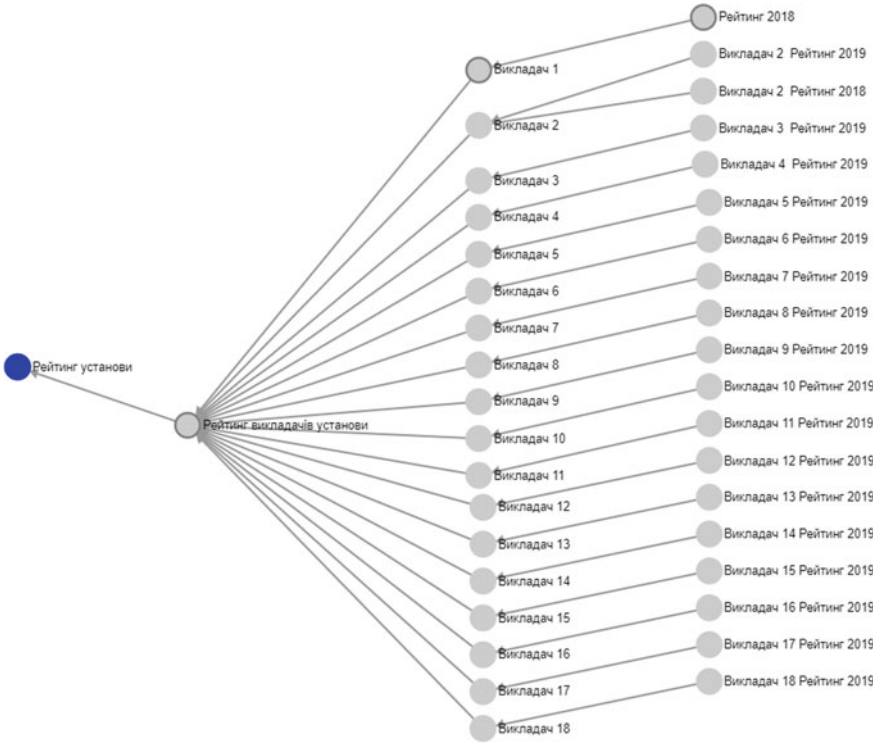



**Fig. 2** Top level ontology

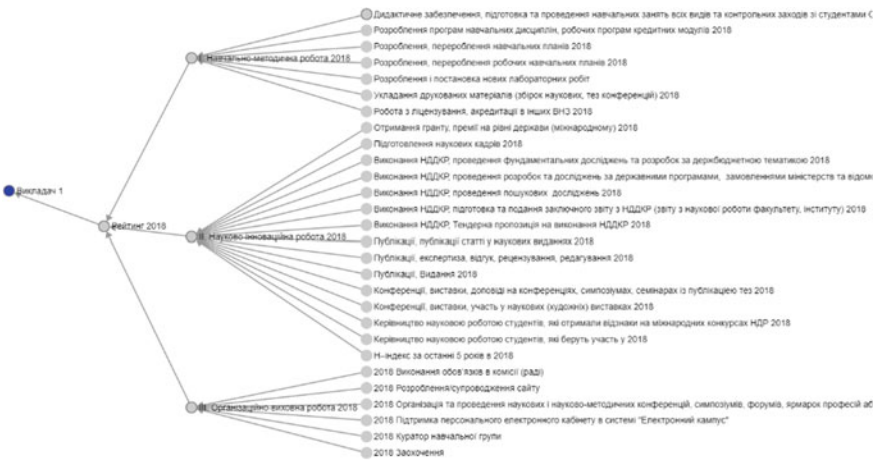



**Fig. 3** Lower-level ontology

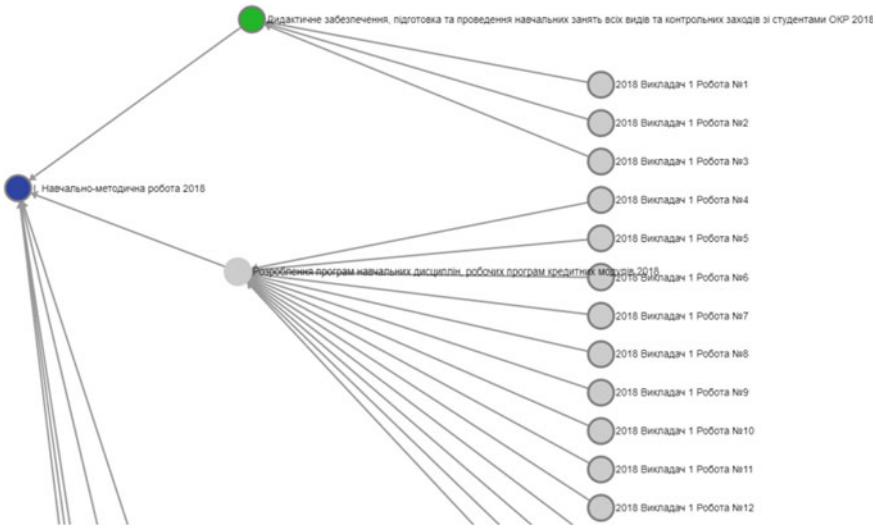



**Fig. 4** Ontology of employee work

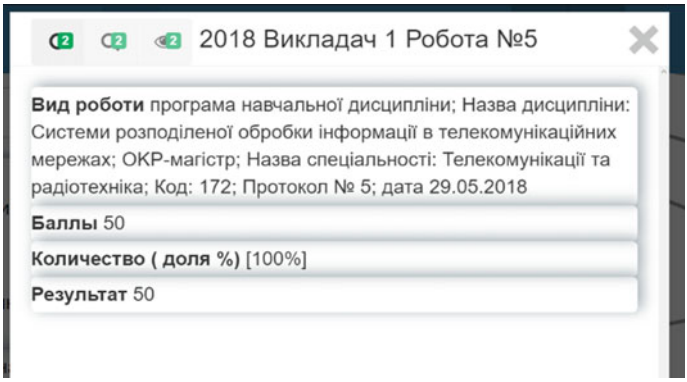



**Fig. 5** Ontology object attribute

In Fig. 7 shows the total number of points for all work performed by the teacher in a given year.

The TEDAOS software environment provides different types of ontology representation. In addition to the graph, it is possible to specify and describe the elements of the ontology in the form of a table. Figure 8 shows the ontological representation of information about employees of the institution in tabular form. The tabular view allows you to integrate ontology information with different editors, allowing you to automatically export and import data.

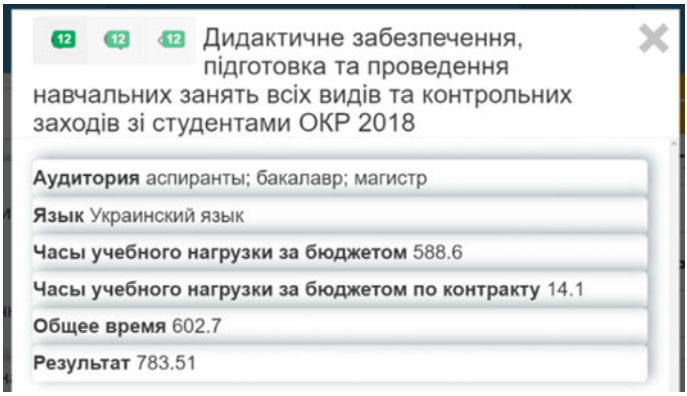



**Fig. 6** Characteristics of works

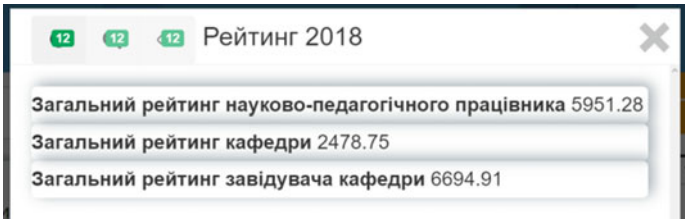



**Fig. 7** Rating the teacher in a given year

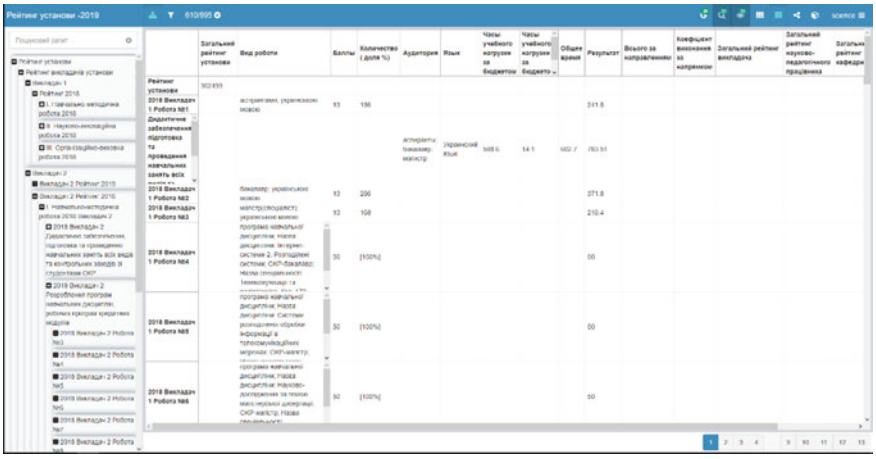


**Fig. 8** Tabular representation of the ontology

Thus, the developed ontological model for assessing the quality of functioning of scientific and educational institutions includes 895 classes, and the total number of elements of the ontology is 2837 elements.

## 6 Conclusions

The paper presents an approach to the representation of information accumulated in various scientific institutions on the basis of ontological model. The ontological representation allows to evaluate the quality of scientific institutions functioning on the basis of national principles for such assessment. The information how to organize the evaluation process as well as generally accepted evaluation criteria are also stored in the ontological model. The associative relations between objects that are determined in ontological model allow to link the institution activities indicators with the evaluation criteria for assessment process automation.

Future researches will focus on further ontology development and its filling with subject domain information, as well as on evaluation of proposed approach usage in comparison with other existing ones.

## References

1. Minsky ML (1979) Frames for representing knowledge. Energy, Moscow
2. Dyakonov VP, Borisov AV (2007) Frame model of knowledge representation. In: Fundamentals of artificial intelligence, p. 30. Smolensk
3. Helbig H (2006) Knowledge representation and the semantics of natural language. Springer, Berlin
4. Poon H, Domingos P (2009) Unsupervised semantic parsing. In: Proceedings of the 2009 conference on empirical methods in natural language processing, Singapore, 6–7 August 2009, pp 1–10
5. Zmitrovich AI (1997) Intelligent information systems. NTOOO "TetraSystems", Minsk, p 378
6. Kruglov VV (2002) Intelligent information systems: computer support for systems of fuzzy logic and fuzzy inference. Publishing House of Physical and Mathematical Literature, p 256
7. Gavrilova TA, Khoroshevsky VF (2000) Knowledge bases of intelligent systems, Textbook. Peter, St. Petersburg
8. Marinicheva MK (2008) 100% knowledge management: a guide for practitioners. Alpina Business Books, Moscow
9. George FL (2009) Artificial intelligence: strategies and methods for solving complex problems, 4th edn. Trans. from English. Williams, Moscow
10. Gruber TR (1993) A translation approach to portable ontologies. Knowl Acquisit 5(2): 199–220. http://www-ksl.stanford.edu/kst/what-is-an-ontology.html
11. Gruber TR (2008) Ontology to appear in the encyclopedia of database systems. In: Liu L, Özsu MT (eds) Springer Verlag. http://tomgruber.org/writing/ontologydefinition-2007.html
12. Soloviev VD, Dobrov BV, Ivanov VV, Lukashevich NV (2006) Ontologies and thesauri: textbook, Kazan, Moscow, p. 176
13. Aldosari B, Alanazi A, Househ M (2017) Pitfalls of ontology in medicine. Stud Health Technol Inf 238:15–18

14. Popova M, Stryzhak O, Mintser O, Novogrudska R (2020) Medical transdisciplinary cluster development for multivariable COVID-19 epidemiological situation modeling. In: 2020 IEEE international conference on bioinformatics and biomedicine (BIBM), pp 1662–1667. https://doi.org/10.1109/BIBM49941.2020.9313204
15. Deambrosis M, Motz R, Eliseo M (2021) UDL ontology: an ontology for education in the diversity. In: Trends and applications in information systems and technologies. https://doi.org/10.1007/978-3-030-72660-7_24
16. Popova M, Rina N (2020) Transdisciplinary system for student youth educational and research activities support. In: 2020 IEEE 15th international conference on computer sciences and information technologies (CSIT), pp 320–323. https://doi.org/10.1109/CSIT49958.2020.9321894
17. Musulin E, Roda F, Basualdo M (2012) Ontology-based expert system for process supervision. Comput Aided Chem Eng 30:957–961. https://doi.org/10.1016/B978-0-444-59520-1.50050-6
18. Littman D (2006) Constructing expert systems as building mental models or toward a cognitive ontology for expert systems. In: Knowledge representation and organization in machine learning. https://doi.org/10.1007/BFb0017218
19. Deshmukh R, Raut B (2021) A comprehensive review on ontology and semantic web. https://doi.org/10.1007/978-981-33-6307-6_29
20. Giri K (2011) Role of ontology in semantic web. DESIDOC J Libr Inf Technol 31:116–120. https://doi.org/10.14429/djlit.31.2.863
21. Karkaletsis V, Fragkou P, Petasis G, Iosif E (2011) Ontology based information extraction from text. In: Paliouras G, Spyropoulos CD, Tsatsaronis G (eds) Knowledge-driven multimedia information extraction and ontology evolution, lecture notes in computer science, vol 6050. Springer, Heidelberg, pp 89–101. https://doi.org/10.1007/978-3-642-20795-2_4
22. Veena J, Raj VP, Shine G (2020) Ontology-based information extraction framework for academic knowledge repository. In: Proceedings of fifth international congress on information and communication technology, ICICT 2020, London, vol 2. https://doi.org/10.1007/978-981-15-5859-7_6
23. Globa L, Novogrudskaya R, Zadoienko B, Stryzhak OY (2020) Ontological model for scientific institutions information representation. In: 2020 IEEE international conference on problems of infocommunications. science and technology (PIC S&T), pp 255–258. https://doi.org/10.1109/PICST51311.2020.9467984
24. Globa L, Novogrudskaya R, Oriekhov O (2013) Method of heterogeneous information resources structuring and systematizing for Internet portals development. In: Eurocon 2013, pp 319–326. https://doi.org/10.1109/EUROCON.2013.6625003
25. Volckmann R (2007) Transdisciplinarity: Basarab Nicolescu Talks with Russ Volckmann. Integral Rev J 4:76
26. Barseghyan AA, Kupriyanov MS, Stepanenko VV, Cold II (2007) Data analysis technologies. Data Mining, Visual Mining, Text Mining, OLAP - 2nd edn, revised and additional. BHV-Petersburg, St. Petersburg, p 384
27. Klein M, Bernstein A (2001) Searching for services on the semantic web using process ontologies. In: The first semantic web working symposium (SWWS-1). Stanford, CA, USA
28. Velychko VY, Popova MA, Prykhodniuk VV, Strizhak OE (2017) TODAOS – IT-platforma formuvannya transdy‘scy‘plinarny‘x informacijny‘x seredovy‘shh / TEDAOS - IT platform for the formation of transdisciplinary information environments. Weap Sys Milit Equip (1):10–19. (in Ukraine)
29. Popova MA, Strizhak OE (2013) Ontologichny‘j interfejs yak zasib predstavlennya informacijny‘x resursiv v GIS-seredovy‘shhi/Ontological interface as a means of presenting information resources in a GIS environment. Sci Not Taurida Nat Univ 26(65):127–135. (in Ukraine)

# Approach to Uniform Platform Development for the Ecology Digital Environment of Ukraine




**Larysa Globa, Stanislav Dovgyi, Oleh Kopiika, and Oleksii Kozlov**

**Abstract** When information system design at the national level, it is necessary to ensure collecting, storing and processing of significant amounts of heterogeneous data accumulated in the repositories of various organizations and departments, circulating in various incompatible systems and not always available for transdisciplinary analysis. To ensure full access to the information, this research proposes the approach to the Uniform Ecology Platform (UEP) development. The platform will create a single point of access to all types of heterogeneous data of the ecology digital environment of Ukraine. The platform should include business process structures, subsystem structures, information structures, and integration structures. The paper proposes five basic principles of the UEP development, namely: uniform information model; common shared telecommunications infrastructure; clearly defined interfaces; independence between business processes and applied subsystems; using of a distributed system with soft links between its components. To implement these principles, the authors suggest a mathematical description of the UEP elements, tools that support the principles of its development and optimization method of successive processes between subsystems of different decomposition-based information communication systems (ICS). As an example, the paper describes global business process formation for the different ICS and their subsystems. The suggested components that are functioning together guarantee full remote access to environmental data stored in different physically distributed systems.

L. Globa (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: lgloba@its.kpi.ua

S. Dovgyi (✉) · O. Kopiika (✉) · O. Kozlov (✉)
Institute of Telecommunications and Global Information Space of NASU, Chokolivsriybulv Blvd. 13, Kyiv 03186, Ukraine
e-mail: s.dovgii@gmail.com

O. Kopiika
e-mail: okopiyka@gmail.com

O. Kozlov
e-mail: alexey.ua84@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_4

**Keywords** Ecology information system · Uniform platform · Information communication systems · System design

## 1 Introduction

The Law of Ukraine "On Basic Principles (Strategy) of State Environmental Policy of Ukraine until 2030" defines the goal of state environmental policy, which is to achieve good environmental status by introducing an ecosystem approach to all areas of socio-economic development of Ukraine to ensure the constitutional right of everyone citizen of Ukraine for a clean and safe environment, the introduction of sustainable use of nature and the preservation and restoration of natural ecosystems [1].

To ensure balanced usage of nature, a system of complex monitoring of the environment state and supervision (control) in the field of environmental protection, management, reproduction and protection of natural resources is needed.

Unfortunately, there are now some researches that are fragmentary in nature and provide only the collection and processing of information on individual components of the environment. Regarding control in the field of environmental protection, there are only administrative mechanisms.

Therefore, the law provides for the development of the ecology digital environment of Ukraine (EDEU).

EDEU should ensure the transition to crime prevention and environmental monitoring, reduce pressure on the business environment, broad public involvement in environmental control through the construction of an effective system for monitoring compliance with environmental legislation, taking into account best practices in the organization of similar systems in European countries.

In this regard, the urgent task is to create a UEP, the approach to development of which will solve the problems of environmental monitoring, supervision (control) in the field of environmental protection, management, reproduction and protection of natural resources, provided under conditions that there are separate systems that provide only the collection and processing of information on individual components of the environment.

The paper has the following structure. Section 2 is devoted to the state of the art and the review of previous research on the complex automated information and analytical systems development. Section 3 considers the problem of EDEU development on the basis of the UEP. Section 4 considers the mathematical description of the UEP EDEU elements. Section 5 considers the implementation of the development principles for the UEP. Section 6 presents the method for optimizing sequential processes between subsystems of different information communication systems (ICS) that have a decomposition nature. Section 7 provides an example of designing a global business process that uses different ICS and their subsystems.

## 2 State of the Art and Background

One of the first systems that provided access and processing of information through the integration of various geographically distributed systems was the Organizational and Technical System (OTS) "Inform-Chernobyl" (developed from 1990 to 1994) [2]. It was developed as a multi-level information and analytical system to support management decisions by the leadership and staff of the Ministry of Chernobyl of Ukraine in all areas defined in the Regulations of the Ministry, to process information from regional and sectoral structures, from selected localities, to generalization of the data processing results of analytical research contained in dozens of databases of various enterprises and organizations.

At the stage of OTS development, the main design decisions provided for three levels, which are related to a set of organizational measures, information flows, software and hardware, the same type of administrative structures.

The central level of the system - "Level of administrative management and decision-making" was intended to support the activities of the staff of the Ministry of Chernobyl of Ukraine;

Subject-analytical level of the system - "Level of scientific and methodological and information support of decision-making" was intended for line Ministries and agencies, research and specialized organizations that collected, accumulated and primary processing of information on the Chernobyl disaster.

The regional level of the system - "The level of initial filling of information and implementation of a set of countermeasures" was intended for units of regional and district administrations, which carried out on-site coordination of work to eliminate the consequences of the accident.

There was a need to develop 48 subsystems, including 13 - the central level, 11 - subject-analytical and 24 - regional. Regional systems were divided into 3 groups according to their complexity. The main element of the system was the expert information block of the top leadership of the Ministry of Chornobyl, connected with the situation centers of the President, the Verkhovna Rada, and the Cabinet of Ministers. But OTS "Inform-Chornobyl" provided only information exchange, and all business processes were performed separately in each subsystem, analysis using all or a group of subsystems required the desining of additional software components through programming by developers, which is not effective.

Further development of this system was continued in the Ministry of Emergencies and Protection of the Population from the Consequences of the Chernobyl Accident (1994–1999) under the new name Organizational and Technological System "MNS-Inform". But, unfortunately, in this system the task of the business processes automated designing, as well as automation of their execution was not solved either.

At the same time, at the request of the Commission of the European Communities (CES) (1994–1999), a computer system for managing the long-term consequences of the Chernobyl accident was designed together with foreign partners [3]. It was proposed to use a decision support system for the management of man-made contaminated areas based on risk analysis using geographic information technologies. The

disadvantage of this project was the use of outdated technologies for collecting processing and transmitting information.

In addition, the following were designed: The system of information and analytical support of public authorities and management commissioned by the National Agency for Informatization (1996–1998) and the State Committee for Communications and Informatization (1999–2001) and other systems. The advantage of these systems was the usage of client-server technology and the integration environment designing that based on the system core [4].

In recent years, the National Telecommunication Infrastructure of Ukraine has been developed and implemented at OJSC Ukrtelecom (2000–2009); Information-analytical system of the budget process support commissioned by the Verkhovna Rada of Ukraine (2007–2011) [5]; National special purpose telecommunication network commissioned by the State Service for Special Communications and Information Protection (2016–2017); Information system for the tactical level of the Armed Forces of Ukraine since 2018 The advantage of these systems was the use of service-oriented architecture. The disadvantage is that the integration of different subsystems and their elements used an information exchange platform, which was configured for each element of integration.

Many researches have been devoted to the integration of various elements of information systems [6], but all of them mostly consider information exchange and providing users with a single portal to obtain the necessary data and solutions, without considering integration processes at the level of automated integrated business processes which uses heterogeneous subsystems and elements of the integrated system as a whole.

In contrast to existing research, this study proposes to consider a broader concept of integration than the exchange of information between modules and subsystems. Therefore, for a complex multidisciplinary information system, business process structures, subsystem structures, information structures and integration structures are considered as a uniform integration mechanism. We call the whole set of processes the uniform information platform.

Automation describes a wide range of technologies that reduce human intervention in processes. Human intervention is reduced by automated determination of decision-making criteria, relationships of subprocesses and related actions through the implementation of these definitions in software [7–12]. Therefore, one of the tasks is the need for mathematical formalization of the existing state of the information and computing environment by means of ontology, which would develop a universal mechanism for access to all available resources of the UEF.

There is a lot of research on some issues of developing mathematical models of the information systems individual elements integration [13–15], which allows to determine the feasibility of developing an ontology of the entire computational space EDEU, which is physically implemented in its UEP. The various user queries should identify the necessary part of the ontological model that will be used in the business processes formation and execution.

This research is devoted to the mathematical model designing for the UEP that would meet the principles of EDEU development in terms of the possibility of the business processes automated integrated formation.

## 3 Problem Definition

The ecology digital environment of Ukraine (EDEU) provides access to environmental information, is a set of information and communication systems (ICS) for various purposes, which already operate in various government agencies. To develop an information-analytical system, there is need first to solve the problem of combining them. From the experience of designing such systems, the architecture of the UEP is proposed. It consists of 4 main levels, which correspond to the physical organization of the system, vertically, and five levels, which implement 5 principles of its development, horizontally (see Fig. 1).

Vertical levels are:

1. The structure of business processes.
2. Structure of subsystems.
3. Information structures.
4. Integration structures.

In the case of ICS development for a telecommunications operator, it is possible to use advanced elements of Frameworx [16]:

- as a structure of business processes it is possible to use a modified business process map Telecom Operations Map (eTOM);

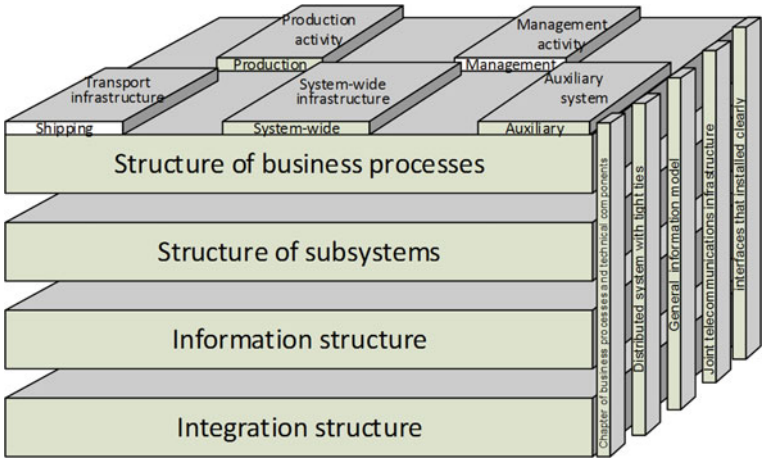



**Fig. 1** Architecture of the UEP

- structure of subsystems - modified map of subsystems Telecom Application Map (TAM);
- information structure - modified data model Shared Information/Data (SID) model;
- integration structure - a modified integration environment - Integration Program (TIP).

The UEP is based on five basic principles, that is: the general information model; common shared telecommunications infrastructure; clearly defined and described in the ontology interfaces; independence of the business processes and applied subsystems; mechanisms for development the distributed system with soft links relations between its components.

All these components in a formalized view form a uniform distributed repository of information and communication resources, the structure of which is described in the ontology form [17–19]. Since the ontology allows to present a set of necessary information for the system operation in the form of a complex graph, it is possible to separate the appropriate subgraph for each business process, to determine a certain required amount of information, software and physical resources. This approach allows to apply a universal mechanism for the business processes formation and implementation, even if there are changes in both information and software, as well as in physical resources. All these changes only need to be specified in the ontological model, editing the relevant information on those information and communication systems where the changes occur.

Thus, the proposed approach to developing the UEP EDEU:

1. Is based on the use of an ontological formalized model that describes the information, software and physical resources.
2. Applies the development principles of the UEP EDEU, which allow integrating an arbitrary set of information and communication systems (ICS) for different purposes.
3. Applies the optimization method for successive processes between the subsystems of different ICS EDEU, which have a decompositional nature.

The effectiveness of the proposed approach is demonstrated by the example of the global business process designing that uses different ICS and their subsystems, when performing tasks in the integrated system EDEU.

## 4 Principles of the UEP Development

The UEP is based on the following principles:

- General information model (X). ICS integration means that subsystems must provide data exchange. To achieve the efficiency of the data exchange process, each subsystem interacts with others through the use of standardized information structures. The Uniform Ecology Platform (UEP) for the data exchanged by

subsystems provides the solution to the problem of converting a hierarchical data model into non-hierarchical models, including a relational data model and vice versa. Such a standardized general information model allows to solve the problems of information exchange in the universal way, which is implemented by the single mechanism, which is a basic component of the UEP.

- General shared telecommunications infrastructure (Q). The era of converged services came in the mid-2000s. Therefore, it is advisable to use a common telecommunications infrastructure (Common Communications Infrastructure, CCI). In this model, Operation Support Systems (OSS) interact with the CCI, not directly with each other. CCI allows subsystems to interact using CCI as the only information repository for their connection. Each subsystem requires only one interface (up to CCI), not many. In this regard, the complexity of the system as a whole is significantly reduced. In addition, the CCI may provide other services, including security services, data conversion, etc. This study proposes to use an intermediate information system that would take over the functions of CCI management, the architecture of such an intermediate information system corresponds to the known models of service delivery platform (SDP) (usually a set of components that provide service delivery architecture, such as service designing, management sessions and protocols).
- Clearly established standardized interfaces (Y). For effective interaction of subsystems with CCI it is necessary to develop standardized interfaces that use modern data exchange technologies. Standardized interfaces are given in the ontological model (for example, Java/JMS or Web services/SOAP), and they also take into account the functionality of the subsystems, the used data, the initial and final conditions, etc. The description of these standardized interfaces should be documented and thus the interfaces become clearly defined and established and considered as additions to the API specifications (Application Programming Interface).
- Independence of business processes and subsystems (Z). When ICS is linked together within the UEP, the business processes supported by them extend to the entire EDEU. As a result, a situation arises when a certain process starts from subsystem A, which processes some data and requires a subsequent call to subsystem B, subsystem B in turn also performs data processing and calls subsystem C, etc. As a result, it is extremely difficult to determine which stage of the process is currently in progress. And even more difficult is the task of changing this process, due to its distributed nature. Assume that the process should be managed as part of a UEP using any mechanism that ensures consistency and is responsible for monitoring the progress of the business process from one subsystem to another. Thus, such a mechanism would initiate a process on subsystem A, which would return control back. After that, this mechanism would call the subsystem B and so on. In this case, it would always be possible to determine which of the business process stages is currently in progress, as control over its progress would already be centralized. In this case, process changes could be processed using certain tools of this mechanism.

- Using a distributed system with smart links between its components (J). The smart relationship between the components assumes that each subsystem is relatively independent of the other subsystems within the overall system. Thus, in an environment with soft links, one subsystem can be modified without affecting others. This principle is seen as enabling plug-and-play subsystems because they are so independent of each other that they can be replaced without affecting the system as a whole. The use of a "distributed system" involves a set of integrated and interacting subsystems, rather than a monolithic subsystem to control all operations [20]. Compliance with the requirements of this principle is achieved through the use of the single ontological model of the entire system as a formalized description of its single information and computing space, i.e. a single information and computing space EDEU.

## 5 Mathematical Description of the UEP Elements

We assume that the UEP consists of the space of sets

$$\langle V, D, F, E \rangle,$$

where: $V$—business processes structure, $D$—subsystems structure, $F$—information structure, $E$—integration structure.

The structure of the *UEP* will be presented as follows:

$$\alpha = \langle v, d, f, e \rangle$$

where $v \subseteq V, d \subseteq D, f \subseteq F$.

A very important element for obtaining new properties is the formalization of redundancies in the structure of the *UEP* through the maximum (+) and minimum (−) sets of parameters:

$$\begin{aligned} v \subseteq v^+, \Delta v = v \backslash v^-, \\ d \subseteq d^+, \Delta d = d \backslash d^-, \\ f \subseteq f^+, \Delta f = f \backslash f^-, \\ e \subseteq e^+, \Delta e = 3 \backslash e^-. \end{aligned}$$

External factors that affect the structure and functionality of the UEP are the development principles, which we will describe as the space of sets

$$\langle X, Q, Y, Z, J \rangle,$$

where: $X$—uniform information model, $Q$—common shared telecommunication infrastructure, $Y$—clearly defined interfaces, $Z$—independence between subsystems and business processes, $J$—distributed system with soft links between its components.

The applicable principles of the EDEU's *UEP* development are described as the rules for the business processes operation:

$$\chi = \langle x, q, y, z, j \rangle,$$

where: $x \subseteq X, |x| \leq |X|, q \subseteq Q, |q| \leq |Q|, y \subseteq Y, |y| \leq |Y|, z \subseteq Z, |z| \leq |Z|, j \subseteq J, |j| \leq |J|$.

Taking into account the set of goals

$$r \subset R,\ r \subset x.$$

Thus, the structure of the *UEP* will be presented as follows:

$$a = v, d, f, e\ :\ \begin{cases} \forall\ w(\alpha, t) \in W,\ |v| < |V| \wedge |d| < |D| \wedge |f| < |F| \wedge |e| < |E|, \\ \Delta v,\ \Delta d,\ \Delta f,\ \Delta e \neq \phi, \\ W(\alpha, r) \rightarrow \max. \end{cases}.$$

The structure of the *UEP* becomes the most effective when four main structural levels are used. The external factors affecting the structure and functionality of the *UEP* are the building principles. Four main structural levels include the following: business processes; subsystems; information structures, and integration structures. Five building principles: uniform information model; common shared telecommunications infrastructure; clearly defined and described in the ontology interfaces; independence between business processes and applied subsystems; building mechanisms for the distributed system with soft links.

EDEU enables the modern infrastructure building based on the convergence of information and communication systems to solve problems of providing access to environmental information, which gives an opportunity to ensure quality and universal access of customers to IT and system services [21–24].

## 6 Optimization Method for Successive Processes Between Subsystems of Different Decomposition-Based ICS

For the synthesis of ICS it is proposed to apply an optimization method for successive processes between the subsystems of different EDEU's decomposition-based information and communication subsystems [25].

Let us consider a universal mechanism for the formation and implementation of business processes. We will demand fulfillment of five information platform

basic principles, namely the use of the following: the uniform information model; common shared telecommunications infrastructure; clearly defined and described interfaces in the ontology; independence between business processes and applied subsystems; building mechanisms for the distributed system with soft links between its components.

Input data:

- $(x_i, y_j)$ —list of ICS subsystems.
- $N$—number of the business process transition stages from one subsystem to another.
- $R$—benefit function of the trajectory of the business process transition from the subsystem $(x_1, y_1)$ to the subsystem $(x_n, y_n)$.
- $r_i^j$—sum of increases in the benefit function of the business process in the subsystems.
- $x$—vector with m coordinates in m-dimensional space, which describes the state of the ICS subsystems of the UEP.
- $y_k$—vector quantity consisting at each stage of certain number of the scalar components $y_{к1}$, $y_{к2}$, …, $y_{кn}$, which is the parameter of the business process transition from one subsystem to another.
- $f_к(x_0)$ —maximum of the benefit functions.

Output data:

- the optimal path of the business process operation between subsystems, which provides for the maximum value of the benefit function.

The method of task solving. Owing to the implementation of the first principle of the uniform information model building, the uniform information model for the data exchange by the ICS subsystems was obtained. When implementing the second principle—the shared telecommunications infrastructure, specifically the network infrastructure is considered as an information system that manages the CCI and provides access to subsystems.

Consequently, the ICS platform was obtained, the state of which is characterized by subsystems $(x, y)$ that get discrete values of $x_i$ $(i = 1, 2, \ldots, n)$ and $y_j$ $(j = 1, 2, \ldots, n)$ (Fig. 2).

When implementing the third principle, while designing clearly established interfaces, the following task is formulated: there is a set of subsystems on the plane, which can determine through the interfaces the business process trajectories from subsystem $(x_1, y_1)$ to subsystem $(x_n, y_n)$. The set of trajectories is formed by the transition from any subsystem to the subsystem that is on the right (left), or to the subsystem that is on top. In total, at the optimal trajectory, $N = 2(n-1)$ stages will be passed or $N = 2(n-1)$ transitions will be made (see Fig. 2).

Using the described interfaces, it is possible for each trajectory of transition from subsystem $(x_1, y_1)$ to subsystem $(x_n, y_n)$ to determine the business process benefit function $R$, which is formed as the sum of increases of benefit functions $r_i^j$ in all subsystems through which business process will pass.

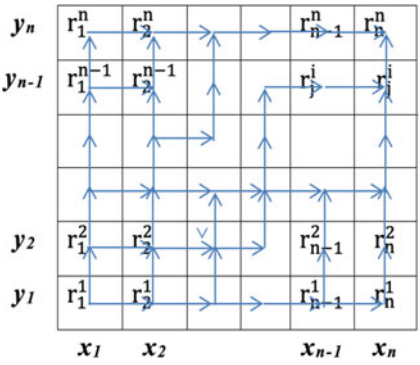



**Fig. 2** Trajectory of the business process transition from the subsystem ($x_1$, $y_1$) to the subsystem ($x_n$, $y_n$)

Thus, the implementation of the fourth principle is simulated—the independence between business processes and applied subsystems. So it's possible to form business processes that use different ICS and their subsystems.

The UEP is described as the state of ICS subsystems, which is characterized by a vector $x$ with $m$ coordinates in $m$-dimensional space. To use global business processes during the stages, the vector $x$ is transformed, starting from its initial value $x_0$, by obtaining another $N-1$ value $x_0, x_1, \ldots, x_{N-1}$. It's assumed that the transition of the vector from one state to another at each stage occurs according to the dependence $x_{k+1} = T(x_k, y_k)$, where $y_k$ is a vector quantity consisting at each stage of a certain number of scalar components $y_{k1}, y_{k2}, \ldots, y_{kn}$, and which is a transition parameter.

For the given vector of the business process state $x_k$ at the beginning of any stage, the state of the subsystems at the end of this stage, or at the beginning of the next one depends on the transformation parameter $y_k$ chosen. Setting all possible values of the transformation parameter at each stage, we obtain all possible variants of the multistage process or all possible ways of the business processes transition from one subsystem $x_0$ to the final $x_k$. The transformation parameter $y_k$ choice at each stage corresponds to the solution of the transition way or the subsystems transformation from state $x_k$ to state $x_{k+i}$. We call the selected sequence of transformation parameters $y_0, y_1, \ldots, y_{N-1}$, as the service of the *i-th* business process.

Let the goal of the multistage transformation process be to achieve the optimum of some scalar quantity $R$, which characterizes the existence of interfaces between subsystems. For certainty, we consider $R$ to be a benefit function and require for a multi-stage process that the benefit function be maximal. At each stage, the benefit function will depend on the system initial state $x_0$ and the selected service $y_0$, $y_1$, $y_{N-1}$, namely

$$R_1 = R_1(x_0,\ y_0);$$
$$R_2 = R_2(x_0,\ y_0,\ y_1);$$
$$R_N = R_N(x_0,\ y_0,\ y_1, \ldots,\ y_{N-1}).$$

In the case of the optimal business process path $y_0, y_1,\ldots, y_{N-1}$, the benefit function will get the maximum value and depend only on the initial state of the system. Denoting the maxima of the benefit functions as $f_{\kappa}(\boldsymbol{x_0})$, we obtain:

$$\begin{aligned} f_1(\boldsymbol{x_0}) &= \max_{\boldsymbol{y_0}} R_1(x_0, y_0); \\ f_2(\boldsymbol{x_0}) &= \max_{\boldsymbol{y_0}, \boldsymbol{y_1}} R_2(x_0, y_0, y_1); \\ f_N(\boldsymbol{x_0}) &= \max_{\boldsymbol{y_0}, \boldsymbol{y_1}, \ldots, \boldsymbol{y_{N-1}}} R_N(x_0, y_0, y_1, \ldots, y_{N-1}). \end{aligned} \tag{1}$$

Indices $y_0$, $y_1$ indicate that the largest value of T will be sought throughout the whole set of subsystems or sequences $y_0$, $y_1,\ldots$, $y_{N-1}$. Direct determination of the expression (1) maximum by blind search is often a practically impossible task. The use of the successive processes optimization method based on the optimality principle, simplifies the determination of the maximum benefit function $\boldsymbol{f}_N(x_0)$ and the formation of the optimal business process path $y_0$, $y_1,\ldots$, $y_{N-1}$, which ensures the benefit function maximum.

According to the principle of optimality, whatever the initial state $x_0$ and the initial solution $y_0$, the business process path $y_0, y_1,\ldots, y_{N-1}$ should be optimal and the benefit function at the N -1 stage should be maximum.

We apply this principle sequentially to *two-*, *three-*, … and *N*-stage processes.

For one stage, as already mentioned, we have

$$f_1(\boldsymbol{x_0}) = \max_{yo} R_1(\boldsymbol{x_0}, \boldsymbol{y_0}).$$

The state of the system in the second stage of the two-stage process

$$x_1 = \boldsymbol{T}(\boldsymbol{x_0}, \boldsymbol{y_0}).$$

Under the condition, whatever $x_0$, $y_0$, the second stage should be optimal, namely

$$f_1(\boldsymbol{x_1}) = \boldsymbol{f_1}\big|\, \boldsymbol{T}(\boldsymbol{x_0}, \boldsymbol{y_0})\big| = \max_{y1} R_1(x_1, y_1).$$

The total benefit function of the two-stage process

$$R_2(x_0, y_0, y_1) = R_1(x_0, y_0) + \boldsymbol{f_1}\big|\, \boldsymbol{T}(\boldsymbol{x_0}, \boldsymbol{y_0})\big|.$$

Accordingly, the maximum value of the benefit function for a two-stage process

$$f_2(\boldsymbol{x_0}) = \max_{yo}\{R_1(x_0, y_0) + \boldsymbol{f_1}\big|\, \boldsymbol{T}(\boldsymbol{x_0}, \boldsymbol{y_0})\big|\}. \tag{2}$$

For the three-stage process, we will similarly have:

$$f_3(\boldsymbol{x_0}) = \max_{yo}\{R_1(x_0, y_0) + \boldsymbol{f_2}\big|\, \boldsymbol{T}(\boldsymbol{x_0}, \boldsymbol{y_0})\big|\}. \tag{3}$$

and finally for *N* stages

$$f_N(\boldsymbol{x_0}) = \max_{yo}\{R_1(x_0, y_0) + \boldsymbol{f_{N-1}}\big|\, \boldsymbol{T}\big(\boldsymbol{x_0}, \boldsymbol{y_0}\big)\big|\}. \quad (4)$$

Recurrent relation (4) enables to sequentially calculate the required value of $f_N(x_0)$ and the optimal solution $y_0$ at the first stage for the $N$-stage process, which will be some function of the system initial state, i.e. $Y_0 = y_N(x_0)$.

Since after determining $Y_{0б}$, $x_1$ is finally defined, to determine $Y_1$ you need to discard the initial business process stage and turn to *(N-1)* stage process with an initial state of $x_1$, namely to define $Y_1$ as a function of $x_1$:

$$Y_1 = y_{N-1}(x_1) = y_{N-1}[T(x_0, Y_0)].$$

In this way, other members of the optimal business process are determined:

$$Y_2 = y_{N-2}[T(x_1, Y_1)]$$
$$Y_{N-1} = y_1[T(x_{N-2}, Y_{N-2})].$$

In order to implement the fifth principle—use of the distributed system with soft links between its components,—the following task is solved.

The use of a "distributed system" involves a set of integrated and interacting subsystems, rather than a monolithic subsystem to manage all operations of the enterprise. The connectivity of the system components is ensured by a uniform information repository of the system elements models, which is one of the components of the UEP, which is built based on the ontological model.

Thus, in an environment with soft links, one subsystem can be modified without affecting other subsystems. In this case, the benefit function $R$ is corrected, and task (1) is solved for the new benefit function $R$.

## 7 Example of the Global Business Process Computation Using Different ICS and Their Subsystems

Input data:

In the case under consideration, we have 5 systems for EDEU, which consist of subsystems set:

- $y_1(x)$—function that characterizes the subsystems for automation of management activities;
- $y_2(x)$—function that characterizes the subsystems for automation of production activities;
- $y_3(x)$—function that characterizes the subsystems for automation of communication infrastructure;
- $y_4(x)$—function that characterizes the subsystems for automation of system software support;

- $y_5(x)$—function that characterizes the subsystems for automation of secondary functions.

Output data:

- The optimal business process path from state $(x_1, y_1)$ to state $(x_n, y_5)$ realized along such a trajectory that the total benefit function was the maximal.

The method of task solving. In this case, the state of the platform is characterized by the stages of the business process $(x, y)$, which get $n$ discrete values $x_i$ $(i = 1, 2,..., 5)$ and $y_j$ $(j = 1, 2,.., 5)$.

We need to design a business process from state $(x_1, y_1)$ to state $(x_n, y_5)$.

There are many processes or trajectories of the business process transition from $(x_1, y_1)$ to $(x_n, y_5)$ on the plane.

Let's assume that for each trajectory of transition from state $(x_1, y_1)$ to state $(x_n, y_5)$ the benefit function $R$ is defined, which is formed as the sum of increases of benefit functions $r_i^j$ in all subsystems through which the business process route will pass.

It is necessary to ensure the data transition from subsystem $(x_1, y_1)$ to subsystem $(x_n, y_n)$ along such a trajectory that the total benefit function was maximal. Thus, the whole transition process of the system from the initial state to the final one is $N = 2(n-1)$-stage process. Provided that, at each stage it is necessary to accept such transition direction (upwards or to the right or to the left) to receive the greatest value of total benefit function $R$.

The task will be solved by the method of successive processes optimization.

The method of successive processes optimization is based on the intuitive principle of optimality, which enables to solve the task of multistage processes optimization by sequentially building recurrent relations. To apply the principle of optimality in this example, we divide the business process path trajectory into two sections: from the beginning to some intermediate $(x_i, y_j)$ stage and from this intermediate stage to the final $(x_n, y_5)$ one.

According to optimality principle, whatever the initial section of the trajectory and whatever the value of the benefit function at stage $(x_i, y_i)$, in order to still get the highest effect in the current situation, the last section of the trajectory from $(x_i, y_i)$ to $(x_n, y_5)$, should be optimal. Obviously, the closer the intermediate stage $(x_i, y_i)$ is to the final one, the easier it is to determine the optimal trajectory.

Thus, for the intermediate stage $(x_{n-1}, y_{n-1})$, the largest value of the benefit function $R_{n-1}^{n-1}$ will be equal to the largest of the two sums:

$$R_{n-1}^{n-1} = \max\begin{pmatrix} R_{n-1}^{n} + r_{n-1}^{n-1} \\ R_{n}^{n-1} + r_{n-1}^{n-1} \end{pmatrix}, \text{where } R_{n-1}^{n} = R_n^n + r_{n-1}^n,\ R_n^{n-1} = R_n^n + r_n^{n-1}.$$

Along with the definition of $R_{n-1}^{n-1}$, the trajectory from $(x_{n-1}, y_{n-1})$ to $(x_n, y_n)$ is determined, so moving to the right or left, or upward. After determining $R_{n-1}^{n-1}$ you can find:

$$R_{n-2}^{n-1} = \max\begin{pmatrix} R_{n-1}^{n-1} + r_{n-2}^{n-1} \\ R_{n-2}^{n} + r_{n-2}^{n-1} \end{pmatrix},$$

**Fig. 3** The sum of increases in the profit function

| | | | $r_j^i$ | | |
|---|---|---|---|---|---|
| $y_5$ | 7 | 3 | 2 | 6 | 12 |
| $y_4$ | 4 | 8 | 9 | 10 | 7 |
| $y_3$ | 1 | 5 | 1 | 5 | 3 |
| $y_2$ | 4 | 9 | 3 | 7 | 6 |
| $y_1$ | 6 | 1 | 9 | 11 | 8 |
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |

**Fig. 4** The benefit function

| | | | $R_j^i$ | | |
|---|---|---|---|---|---|
| $y_5$ | 30 | 23 | 20 | 19 | 12 |
| $y_4$ | 50 | 46 | 38 | 29 | 19 |
| $y_3$ | 52 | 51 | 39 | 34 | 22 |
| $y_2$ | 64 | 60 | 44 | 41 | 36 |
| $y_1$ | 70 | 62 | 51 | 52 | 36 |
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |

$$\mathrm{R}_{\mathrm{n}-1}^{\mathrm{n}-2} = \max\begin{pmatrix} \mathrm{R}_{\mathrm{n}-1}^{\mathrm{n}-1} + \mathrm{r}_{\mathrm{n}-1}^{\mathrm{n}-2} \\ \mathrm{R}_{\mathrm{n}}^{\mathrm{n}-2} + \mathrm{r}_{\mathrm{n}-1}^{\mathrm{n}-2} \end{pmatrix},$$

where $R_{n-2}^n = R_{n-1}^n + r_{n-2}^n$, $R_n^{n-2} = R_n^{n-1} + r_n^{n-2}$.

Further, it is similarly possible to determine the highest value of the benefit function $R_j^i$ for any intermediate stage $(x_i, y_i)$ when transition is realized from this intermediate stage to the final one. Simultaneously with the defining $R_j^i$, the transition trajectory from $(x_1, y_1)$ to the position $(x_n, y_n)$ is determined, which is the solution of the task.

Figure 3 shows the numerical values of $r_j^i$ and Fig. 4 shows the numerical values of $R_j^i$ and the optimal trajectory from $(x_1, y_1)$ to $(x_5, y_5)$. This example illustrates the main idea of the method.

Thus, using the example of five systems EDEU, consisting of the subsystems set, demonstrates the business process designing, which is based on the method of sequential processes optimization. Such an approach allows to consistently solve problems of multi-stage business processes optimization by developing recurrent relationships.

In practice, this approach enables to design a universal mechanism for the business processes for EDEU formation and implementation, taking into account five basic principles of the information platform, namely using: a uniform information model; common shared telecommunications infrastructure; clearly defined and described in the ontology interfaces; independence between business processes and applied subsystems; building mechanisms for the distributed system with soft links between its components.

Thus, the proposed approach to building the UEP for the National Automated Ecological Information and Analytical System enables to increase its flexibility,

have access to the necessary information and computing resources at any time, and minimizes access time.

## 8 Conclusions

1. The National Automated Ecological Information and Analytical System providing access to environmental information is a set of various purpose information and communication systems.
2. The paper presents the solution to the task for EDEU information and communication systems integration based on the UEP.
3. The EDEU's UEP is formed based on a common ontological model and shall include:
   - business process structures;
   - subsystem structures;
   - information structures;
   - integration structures.
4. The EDEU's UEP shall be based on five basic principles, namely:
   - uniform information model;
   - common shared telecommunications infrastructure;
   - clearly defined interfaces;
   - independence between business processes and applied subsystems;
   - use if the distributed system with soft links between its components.
5. The EDEU UEP was built using the following resources:
   - mathematical model of system architecture as a whole in the form of ontology elements;
   - five principles of the UEP development;
   - method of successive processes optimization between subsystems of different decomposition-based ICS.

The discussed cumulative results represent the main conceptual provisions of the UEP for the National Automated Ecological Information and Analytical System that provides access to ecological information. They allow increasing the flexibility of the system as a whole, warrant access to necessary information and computing resources at any time, and minimize access time.

Further research will be needed to improve the ontological model representing all aspects of the system components description, develop and implement the distributed repository of the uniform information environment of the system, and elaborate specific technical and technological requirements for the EDEU.

## References


1. The Law of Ukraine «On Basic Principles (Strategy) of State Environmental Policy of Ukraine until 2030» (2019). https://zakon.rada.gov.ua/laws/show/2697-19#Text
2. Dovgiy S, Kopiika O (2001) Automated system to support decision-making in the aftermath of the Chernobyl accident, Informatization of Aerospace Geology, pp 211–266. Scientific thought, Kyiv
3. Lochard J, Belyaev S (1996) Decision aiding system for the management of post-accidental situations. In: Lochard J, Belyaev S (eds) Final Report, Joint Study Project no 2, European Commission, DG XII, Brussels: EUR 16534 EN
4. Dovgiy S, Kopiika O, Cherepin Y (2004) Principles of regional informatization. VPC «TYRAZH», Kyiv
5. Dovgiy S, Serhiyenko I (2013) Information and analytical support of the budget process. Information systems, Kyev
6. Serhiyenko I, Stetsyuk P, Koshlay L (2009) Models and information technologies for decision support during structural and technological changes. Cybern Syst Anal (2):26–49
7. Groover M (2014) Fundamentals of Modern Manufacturing: Materials, Processes, and Systems, Wiley, Hoboken
8. Bahrin MA, Othman MF, Nor Azli NH, Talib MF (2016) Industry 4.0: a review on industrial automation and robotic. J Teknol 78:6–13
9. Lars A et al (2016) German Standardization Roadmap: Industry 4.0, Version 2. DIN e.V., Berlin
10. Schlaepfer RC, Koch M (2015) Industry 4.0: challenges and solutions for the digital transformation and use of exponential technologies. Creative Studio at Deloitte, Zurich, Deloitte AG, p 32
11. Willliam MD (2014) Industrie 4.0–Smart Manufacturing For The Future. Germany Trade & Invest, Berlin
12. Kagermann H, Wahlster W, Helbig J (2013) Recommendations for Implementing yhe Strategic Initiative Industrie 4.0: Final Report of the Industrie 4.0 Working Group, Ulrike Findeklee: Acatech–National Academy of Science and Engineering
13. Sviridov S, Kuryan A (2003) IDEF0: Functional modeling of processes, IP «Oriyentsoft». https://www.orientsoft.by/pdf/IDEFO_FM.pdf
14. Steklov V, Berkman L, Kilchitsky E (2004) Optimization and modeling of communication devices and systems, Textbook. Machinery, Kyiv
15. Choi M-J, Ju H-T, Hong JW-K, Yun D-S (2008) Design and implementation of web services-based NGOSS technology specific architecture. Ann Telecommun Spec Issue Next Gener Netw Serv Manag 63(3–4):195–206
16. Collection of knowledge on business process management: BPM 3.0 = BPM Version 3.0. Guide to the Business Process Management Common Body of Knowledge, Alpina Publisher, Moscow (2016)
17. Globa LS, Gvozdetska NA, Novogrudska RL (2021) Ontological model for data processing organization in information and communication networks, System Research and Information Technologiesthis link is disabled, pp 47–60
18. Popova M, Globa L, Novogrudska R (2021) Multilevel ontologies for big data analysis and processing. In: Proceedings of International Conference on Applied Innovation in ITthis Link is Disabled, vol 9, no 1, pp 41–53
19. Globa L, Novogrudska R, Koval A (2018) Ontology model of telecom operator big data, 2018/6/4. In: 2018 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp 1–5
20. Kopiika O, Skladannyi P (2021) Use of service-oriented information technology to solve problems of sustainable environmental management. In: CEUR Workshop Proceedings, 2021, 3021, pp 66–75
21. Jonathan J (2010) BICSI Data Center Standard: A Resource for Today's Data Center Operators and Designers, Jew Jonathan, BICSI News Magazine, May/June 2010

22. Niles S (2011) Standardization and Modularity in Data Center Physical Infrastructure, Niles Susan, Schneider Electric
23. Telecommunications Infrastructure Standard for Data Centers, TIA STANDARD TIA-942, Telecommunications Industry Association, April 2005
24. Data Center Design and Implementation Best Practices, ANSI/BICSI 002-2011, Committee Approval, January 2011, First Published: March 2011
25. Globa L, Dovgiy S, Kopiika O, Kozlov O (2021) Approach to building uniform information platform for the national automated ecological information and analytical system. In: CEUR Workshop Proceedings, 3021, pp 53–65

# Cloud-Based Technologies for Data Processing in Ukraine: International Context




**Andrii Shelestov, Bohdan Yailymov, Hanna Yailymova, Svitlana Nosok, and Oleh Piven**

**Abstract** During last time we have faced with big data problem in Earth observation domain. Fortunately, cloud solutions such as Amazon Web Services, Google Earth Engine and others platforms provide an access to Sentinel-1, Sentinel-2 and Landsat data with spatial resolution from 10 to 30 m, opportunities for quick and convenient way of geospatial data processing and usage for a lot of different information products retrieval like land cover classification maps, crop state monitoring etc. During last few years we had experience of cloud-based technologies usage within scientific and innovation projects, which are supported by European Commission, World Bank, United Nations Development Program, GEO Committee and have experience of open-source software development and machine (deep) learning usage in cloud environment, namely Open Data Cube. This package provides the opportunities for data collection, deployment and provision on the base of 3D model of data representation. The developed technologies are implemented on diverse cloud platforms and solve various types of applied problems, in particular, monitoring of agricultural lands, assessment of sustainable development indicators at the national level. All of these questions will be described in our chapter in more detail.

B. Yailymov (✉)
Space Research Institute National Academy of Sciences of Ukraine and State Space Agency of Ukraine, Glushkov Avenue 40, 4/1, Kyiv 03187, Ukraine
e-mail: yailymov@gmail.com

A. Shelestov (✉) · H. Yailymova (✉) · S. Nosok (✉) · O. Piven (✉)
Institute of Physics and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: andrii.shelestov@gmail.com

H. Yailymova
e-mail: anna.yailymova@gmail.com

S. Nosok
e-mail: nos.sv.ol@gmail.com

O. Piven
e-mail: pivolegbor@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_5

**Keywords** Cloud-based information technology · Agricultural monitoring · Time series data analysis · Satellite monitoring

## 1 Introduction

The expert group, which is consists of specialists from Department of Mathematical Modelling and Data Analysis in Educational and Research Institute of Physics and Technology of NTUU "Igor Sikorsky Kyiv Polytechnic Institute" and Department of Space Information Technologies and Systems of Space Research Institute NAS of Ukraine an SSA Ukraine (SRI NASU-SSAU), is the member of JECAM and GEOGLAM programs. The experts participate in different international and national projects and have been providing satellite monitoring products and services for more than 10 years to the government and various organizations including activities under the umbrella of Group on Earth Observation Committee (GEO). The main goal within all of these programs and projects is to influence the digitalization of the economy and decision-making processes of our state with satellite monitoring using.

Over the past five years, Ukrainian agricultural industry has become one of the main sectors of growth in Ukrainian economy. Today, a vast majority of export revenue is generated by agricultural sector. This industry also helps to develop related and complementary industries, at the same improving the country's international reputation. At the same time, the level of digitalization in Ukraine is much lower, than in the European Union, and agricultural policy support lacks sophisticated management tools, such as the EU's Integrated Administration and Control System, which includes a Land Parcel Identification System (LPIS). This is mostly due to the low level of modern technologies utilization by businesses and state entities. Therefore, in this paper we consider the steps taken by scientists on the way to digitalization of land resources of Ukraine using cloud technologies.

In July 2021, Ukraine opened the land market. To ensure transparency, equity and reliability of this process, objective information on real land use and its history, crop development state in each vegetation season is required. All these indicators significantly affect the market value of the land. Within the World Bank program "Supporting Transparent Land Governance in Ukraine" with financial support from the EU, which addresses these issues, we carried out satellite monitoring of land cover/land use (LC/LU) in Ukraine, analyzed the feasibility of Google Earth Engine cloud-based technology for processing large amount of data to analyze crop state using open and free Sentinel-1/2 satellite data [1].

Another area of activities that all developed countries are currently engaged in is the Sustainable Development Goals (SDGs) assessment. The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, makes an emphasis on the SDGs and corresponding indicators that characterize complex interactions of human influence and environment state. Many indicators are based on geospatial information, and can be derived from satellite data integrated

along with in-situ measurements and models, including weather forecasts, biophysical estimation and classification (machine learning) models. Therefore, there is a need to develop methods and tools that will allow derivation of SDGs indicators as geoinformation products [2].

Nowadays, methodologies for calculation SDGs indicators are mainly based on coarse spatial resolution (300 m) satellite data and global products. However, our previous studies for Ukraine showed that global coarse resolution products are not very accurate at local (country) scale, especially for agriculture potential estimation, due to producing mixtures in pixel values from combining different land cover types in a single pixel [3]. That is why within the framework of the joint grant from Amazon and the GEO Committee GEO-Amazon "Methodology for SDGs indicators assessment" the authors developed an automated technology for SDGs indicators calculation on the cloud platform Amazon Web Services using modern Open Data Cube technologies [4]. To assess agricultural area under productive and sustainable agriculture, we propose to combine this approach with biophysical model WOFOST [5] as an alternative source of information which allows us to increase the temporal resolution of crop state indicators.

Within the international GEO-GEE (Group on Earth Observation and Google Earth Engine) project "Satellite Monitoring for Sustainable Land Management and Agriculture in Ukraine with Google Earth Engine" authors solved the following applied tasks and support decision makers with the objective satellite-based information on the cloud platform: crop area estimation and real land use monitoring at the country level; land cover change monitoring; double crops detection and monitoring; detection and monitoring of crop rotation violations; drought detection and monitoring; SDGs indicators 15.3.1 and 2.4.1 monitoring. Some of the most valuable projects are described shortly below.

World Bank project. Within the projects "Supporting Transparent Land Governance in Ukraine" (2016–2019, 2021) and "Satellite Imagery Processing for Crop Cover Analysis", which is implemented through the EU-funded project "Support to Agriculture and Food Policy Implementation (SAFPI)" and "Moving Forward Together" (2020), we analyzed the feasibility of using Google Earth Engine cloud-based technology to monitor land use, and opportunities for development and usage of solution to analyze the stage of crops worldwide, as well as develop LC/LU maps. This analysis is performed with the use of innovative cloud-based free and open satellite information processing technologies (Sentinel-1 and Sentinel-2), and artificial intelligence models, which are used for building classification maps. Within this project we used random forest classifier (RF) in Google Earth Engine Platform – one of the most popular machines learning methods, consisting in application of the ensemble of decision trees and applied for the tasks of classification and regression. In particular, within the project authors proposed to use a previously developed approach for high resolution LC/LU mapping at country level. The proposed approach outperforms global products in terms of accuracy and spatial resolution.

Amazon-GEO project "Methodology for SDGs indicators assessment". Within the Amazon-GEO project "Methodology for SDGs indicators assessment" (2019–2022) Ukrainian team developed automated workflows for calculation SDGs Indicators grouped by land cover classes. The main goal is to adopt, improve and apply already proposed methodologies, which were used for generating global products with coarse spatial resolution data, to higher spatial resolution data (up to 10 m), which will be better suited for regional products and applications. The following SDGs indicators were investigated within the project: 2.4.1 "Proportion of agricultural area under productive and sustainable agriculture" [6], 11.3.1 "Ratio of land consumption rate to population growth rate" [7], 15.1.1 "Forest area as proportion of total land area" [8], 15.3.1 "Proportion of land that is degraded over total land area" [9]. The informational technology for these indicators was implemented it in the AWS cloud environment. It was based on Open Data Cube technology and included deep learning algorithms for LC/LU classification, biophysical modeling, weather modeling, and satellite data analysis. The technology is scalable and usable for any another country. The main innovation of the project is concerned with the improvements of existing workflows for SDGs indicators calculation by the use of high spatial resolution data and filling gaps between existing global products and national ones. In particular, in [10] propose to use a previously developed neural network approach for high resolution LC/LU mapping at country level. The proposed approach outperforms global products in terms of accuracy and spatial resolution. To assess agricultural area under productive and sustainable agriculture, we propose to combine this approach with biophysical model WOFOST as an alternative source of information which allows us to increase the temporal resolution of crop state indicators. As a case study all these indicators calculated for Ukraine and will be calculate for Argentina and India.

GEO-GEE project "Satellite Monitoring for Sustainable Land Management and Agriculture in Ukraine with Google Earth Engine". The project "Satellite Monitoring for Sustainable Land Management and Agriculture in Ukraine with Google Earth Engine" (2020–2022) was nominated by the Joint Google Earth Engine and GEO Committee Program. Within the project the consortia participants have gone the opportunities of licensed access to conduct scientific research in the Google Cloud Platform. The open-source Google Earth Engine platform provides access to traditional pixel-based machine learning and classification methods. Nowadays deep learning methods outperform traditional machine learning methods on a lot of applied tasks, such as human language translation, computer vision problems or speech recognition and remote sensing data analysis is not an exception. However, there are difficulties and challenges in applying deep learning methods for remote sensing data. One particular issue is on-time parcel classification, which requires extensive use of multi-sensor time-series spatial data. Furthermore, this task critically hinges on the integration of spatial information (e.g., parcel segmentation) and assimilation of external drivers (e.g., weather information, agronomic knowledge on practices). This amount of labeled ground truth data for learning and validation tasks. At the moment such data tend to be scarce and don't have one universal form, especially at the level of large production zones or country level. Thus, within the

international GEO-GEE project Ukrainian team developed operational sustainable technology based on deep learning methods for LC/LU classification based on satellite data for governmental use and decision making. The main goal was to improve existing deep learning methods for precise LC/LU mapping using satellite data (up to 10 m) that have been previously developed and validated within the World Bank program. Within the project, we have adapted and modified the deep learning methods for LC/LU classification at the country level.

## 2 Data

### *2.1 Satellite Data*

There are several sources of free satellite data available on regular basis and that can be used for LC/LU mapping. In particular, we use time-series of Sentinel-1/SAR data acquired from 1st of April to 15th October for the research year and Sentinel-2 optical composites [11] for the same period. This temporal period is selected to match a crop calendar for Ukraine, when the most crops are grown from April to October. Input satellite data format is hardly related to used classification method. Sentinel-1 provides nearly 3 Tb of images for Ukraine and Sentinel-2 has about 4 Tb to solve classification task for the identical territory. At the same time, larger countries need more memory for the implementation of the algorithm: from 5 to 6 Tb for Sentinel-1 and from 10 to 12 Tb for Sentinel-2 for one vegetation year. These huge amounts of satellite data are already stored in the Google Earth Engine Platform and Google Cloud Platform and there is no need for their downloading, processing and transferring for further use. If we use the Amazon platform, the following steps of SAR satellite data pre-processing are need: reading the Sentinel-1 GRD archive; correction of coordinates in orbit (apply orbit file); specl-filtration (Lee filter $5 \times 5$); radiometric calibration with conversion to the values of the inverse scatter coefficient of the signal Sigma0 (calibration); the Range-Doppler terrain correction procedure is performed using the given digital terrain model (SRTM with 90 m resolution); data transfer in decibel (LinearToFromdB); creating a Data Stack (CreateStack); saving imagery bands (VV, VH); merging of bands VV and VH within a single granule. Atmospheric correction with Sen2Cor algorithms and clouds and shadows masking are used for optical data. In the Table 1 are shown the dates that were used for classification mapping for Ukraine in 2021. Data coverage for Ukraine is shown in Fig. 1.

**Table 1** Used satellite data for Ukraine (as an example for 2021)

| Satellite data | Dates |
|---|---|
| Sentinel-2 (composites) | From 01.04.2021 to 31.05.2021;<br>From 01.05.2021 to 16.07.2021;<br>From 07.07.2021 to 08.08.2021;<br>From 08.08.2021 to 02.09.2021;<br>From 20.08.2021 to 15.10.2021 |
| Sentinel-1 | From 2021–04–01 to 2021–10–15<br>(with revisit time 6 days for Ukraine) |

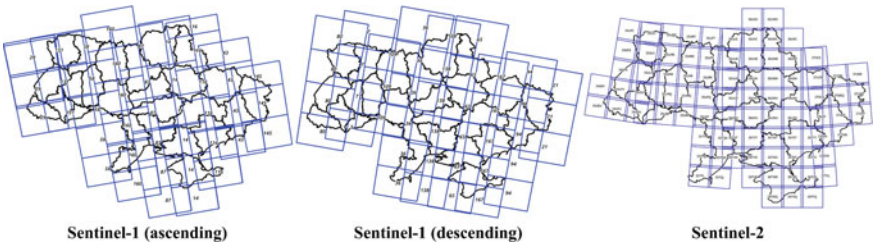



**Fig. 1** Satellite data coverage for Ukraine used for classification mapping

## *2.2 In-Situ Data*

In-situ data collection is carried out annually during the active stage of crop growing to obtain reliable information for neural network training and LC/LU classification map creation. The routes planning of ground survey based on analysis of time series satellite data and official statistical data [12]. The route for in-situ data collection is focused on the accumulation of the maximum number of crops, as well as their maximum diversity. For the data collection, the roads through which you can reach each of the fields were considered. In-situ data collection performed according to JECAM guidelines for in-situ data collection provided in "JECAM Guidelines for cropland and crop type definition and field data collection" at the source [13]. In particular, provided in section III, "Field data collection for cropland and crop type" recommendations were used for in-situ data collection campaign design. According to these recommendations 2 routes were laid for field data collection (on the example of 2021). One of them for winter crops data collection (Fig. 2, green line), and another for summer crops (Fig. 2, orange line). They uniformly cover the territory of Ukraine taking into account stratification (agricultural areas) and the availability of the roads. The route for winter crops in 2021 is about 1400 km, for the summer - about 5500 km.

The data collected after the pre-processing stage (digitization and anomaly detection) are divided into two independent parts for each oblast of Ukraine and for each class – for neural network training and for the resulting classification map testing. The corresponding distribution by classes and oblasts is shown in Table 2.




**Fig. 2** Winter and summer in-situ data collection roads in 2021

**Table 2** In-situ data distribution between test and train sets in 2021 (the color of each cell corresponds to the class on the classification map)

| LC Type (crops) | Train | Test | LC Type | Train | Test |
|---|---|---|---|---|---|
| Cereal crops | 1858 | 1855 | Artificial | 222 | 223 |
| Rapeseed | 241 | 241 | Forest | 223 | 224 |
| Buckwheat | 38 | 38 | Grassland | 888 | 866 |
| Maize | 915 | 914 | Bare land | 135 | 115 |
| Sugar beet | 39 | 39 | Water | 351 | 327 |
| Sunflower | 993 | 992 | Wetland | 107 | 114 |
| Soybeans | 409 | 409 | Alfalfa | 45 | 34 |
| Other crops | 214 | 205 | Gardens | 191 | 176 |
| Peas | 63 | 63 | Grape | 231 | 224 |
| **Total** | **4770** | **4756** | **Total** | **2393** | **2303** |

# 3 Methodology

## *3.1 Classification*

Many global geospatial products are based on land cover classification results and geospatial crop type classification products. Therefore, it is very important to get such classification products on time and with high accuracy at national level. In this regard, the main idea of all our project, which are implemented within cloud platform, is to propose better LC/LU classification approach for Ukraine. The proposed methodology is based on a deep learning approach for LC/LU mapping, in particular an ensemble of neural networks [10, 14]. A committee of neural networks is used for providing crop classification and land cover maps for the territory of interest (starting from whole Ukraine use-case) using high resolution Sentinel-1 and Sentinel-2 imagery and appropriate in-situ data. Time series of satellite data for vegetation period allow achieve better accuracy of land cover classification and could help us more precise discriminate crops and other land cover types. But utilizing time series of high-resolution satellite imagery and it's preprocessing is time consuming task. Another issue is training of deep learning models on time series of satellite data.

To address aforementioned big geospatial data challenges two main powerful cloud platforms are available at the moment: Amazon and Google. We have previous experience with utilizing powerful cloud platforms for LC/LU classification such as Amazon Web Service (AWS) [15] and freely available Google Earth Engine (GEE) [16]. Cloud platforms allow us to overcome challenges of satellite data download and processing. Google Earth Engine platform provides built-in functions and intrinsically-parallel computational access. At the same time, Amazon cloud platform provides an opportunity for exploiting any software for image processing and libraries with advanced classifiers. As well as Amazon does not provide ready to use functionality of satellite data processing and classification, it's utilization requires much more efforts. We have developed an automated workflow for cloud-based crop classification using AWS based on Sentinel-1 and Sentinel-2 imagery. This methodology has been applied for providing 10 m resolution crop classification maps for the Ukraine territory in 2016–2021. This improvement provides much better accuracy in indicators estimation.

## *3.2 SDGs Indicator Assessment*

As follows from documents of United Nations and Paris Summit (2015) another important task is society sustainable development provision. Such a results can be estimated on the base of geospatial information usage. Within this global framework on the base of already developed LC/LU classification at country level, Ukrainian team proposed a unified workflow (Fig. 3) for calculation of SDGs indicators, namely 11.3.1, 15.1.1, 15.3.1 and 2.4.1. [15, 18]. Proposed approach outperforms global

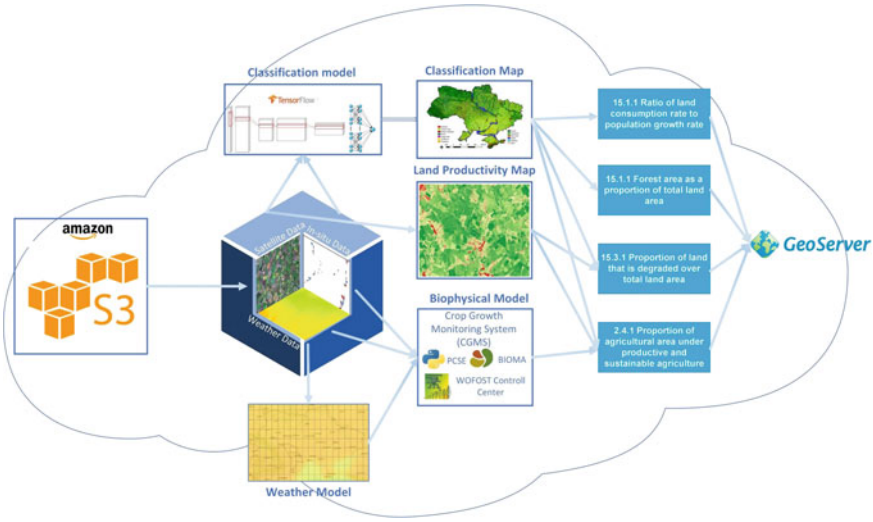



**Fig. 3** Workflow for calculating sustainable development goals indicators 11.3.1, 15.1.1, 15.3.1 and 2.4.1 using cloud data cube technology

products in terms of accuracy and spatial resolution. Often for SDG estimation should be applied NEXUS approach [22].

The assessment of the all indicators is based on classification map and auxiliary geospatial products and data integration procedures. Some of them will be described below in this subsection.

**SDGs Indicator 15.1.1 "Forest Area as Proportion of Total Land Area".** The SDGs indicator 15.1.1 "Forest area as proportion of total land area" is a proportion of forest areas to total land area [19]. According to existing methodology FAO has been collecting and analyzing data on forest area since 1946 [28]. It was done at intervals of 5–10 years as part of the Global Forest Resources Assessment (FRA). FRA 2015 product contains information for 234 countries and territories on more than 100 variables related to the extent of forests, their conditions, uses and values for five points in time: 1990, 2000, 2005, 2010 and 2015. Assessment of forest area is carried out at infrequent intervals in many countries. Access to remote sensing imagery has improved in recent years.

Another globally available product is Global Forest Change produced by University of Maryland with use of time-series analysis of Landsat images in characterizing global forest extent and change from 2000 through 2017 at 30 m resolution.

So, the main required information for indicators is time series of LC/LU maps over the territory of interest. Therefore, we propose to use our own national product with 10-m LC/LU classification, which is obtained by its developed technology.

**SDGs Indicator 15.3.1 "Proportion of Land That is Degraded Over Total Land Area".** The SDGs indicator 15.3.1 "Proportion of land that is degraded over total

land area" [17] is a proportion of degraded land to the total area of the country and based on the analysis of available data or developed at the national level products (namely, Trends in Land Cover, Land Productivity and Carbon Stocks (Table 3) [19, 20]. This indicator is based on statistical principal "One Out, All Out" on evaluation of changes in the sub-indicators. This principle means that we have three types of changes in the sub-indicators, which are depicted as positive or improving, negative or declining and sustainable or unchanging. This methodology unfortunately can't be used in full mode because it is very hard to obtain Carbon Stocks maps for every year at national level [21]. The newest Carbon Stocks map is dated by 2016 year and already outdated as well as the newest open access land cover map dated by 2015.

If one of the sub-indicators has negative changes for some area, then this area has negative productivity. According to existent methodology [22] as negative changes considered following transitions: decrease of carbon stock level over the period of time, decline in land productivity or negative land cover changes (i.e., forest → grassland, forest → cropland, any green area → urban) etc.

Actually, the approach implemented by JRC takes into account total green vegetation and non-suitable for estimation of cropland productivity (the most productive regions often partially covered by forests) [23]. Furthermore, coarse resolution of this global product produces mixtures in pixel values from combining different land cover types in single pixel. That is why we need national Land Cover product with high resolution.

**Table 3** Used datasets description

| Dataset type | Source | Specs | Comments |
|---|---|---|---|
| Land Cover | ESA CCI land cover dataset | 300 m resolution, 1992–2015, 22 classes | Rescaled from 22 classes to 6 mains (Forest, Grassland, Cropland, Artificial, Wetland and Water, Bare land). Basic period 2000–2015 (UNCCD) |
| Land Productivity Dynamics | JRC Productivity Dynamics Dataset | 1 km resolution, 5 classes | Based on time series of SPOT-Vegetation data collected during 1999–2013. 5 qualitative classes of land productivity trends are available (Declining productivity, Early signs of decline, Stable, but stressed, Stable, not stressed, Increasing productivity) |
| Soil Organic Carbon stocks | ISRIC's SoilGrids 250 m | 250 m | Topsoil SOC values (0–30 cm) |

**SDGs Indicator 11.3.1 "Ratio of Land Consumption Rate to Population Growth Rate".** The SDG indicator 11.3.1 "Ratio of land consumption rate to population growth rate" [25, 26] can be calculated using detailed land cover maps built on moderate and high-resolution images for two years at least to estimate ratio of land consumption rate and requires open access statistics about city population. Population statistics over the city provided by statistical authorities is used for population growth rate assessment. In case when national data are not available, it is possible also to use open data such as JRC Global Human Settlement layer [27].

This indicator can be used in every country in the world with use of global coarse resolution LC. Information can be obtained at two levels. First level is city level indicator, which measure sustainability in terms of city and population growing. The second level is country based, which aggregate city's-based indicators within the country and indicate urban area and population growth rate for this country.

**SDGs Indicator 2.4.1 "Proportion of Agricultural Area Under Productive and sustainable Agriculture".** This workflow of SDGs indicator 2.4.1 "Proportion of agricultural area under productive and sustainable agriculture" [21] calculation includes the definition of land productivity in particular over cropland. The methodology implemented by JRC takes into account total green vegetation and non-suitable for estimation of cropland productivity (the most productive regions often covered with forests) [23]. Furthermore, coarse resolution of this global product produces mixtures in pixel values from combining different land cover types in single pixel.

The SDG indicator 2.4.1 could be calculated using the same methodology as proposed for indicator 15.3.1 calculation. This indicator is a proportion of agricultural area that has a positive productivity trend value to the total agricultural area by the rule "One Out, All Out". For this indicator, the same sub-indicators as for indicator 15.3.1 are used, but the area of interest is not the whole area of the country, but rather the agricultural land (cropland) subsetted with use of LC map. For this indicator the use of high spatial resolution satellite images is particularly important, since mixed pixels greatly affect the value of sub-indicator changes.

## 4 Results

As mentioned above, the main prerequisite product for almost all SDG indicators assessment is classification map. Due to this it is very important to estimate classification accuracy. using developed approach for the major crops for Ukraine in 2021 (Fig. 4) the overall accuracy was higher than 94%, that is on the cutting-edge of latest advance in ML technologies. The confusion matrices for classes based on the test independent in-situ data for the map obtained by the Random Forest in Google Earth Engine and Neural Network (NN) approach within the Amazon Web Service environment are shown on Fig. 5 and Fig. 6. Using the Random Forest method overall accuracy was 94.8% and overall accuracy of NN method was 96.7% [24].

Both classification maps (GEE-based and NN) were compared with official statistics, obtained from the State Statistics Service of Ukraine for different crops. For

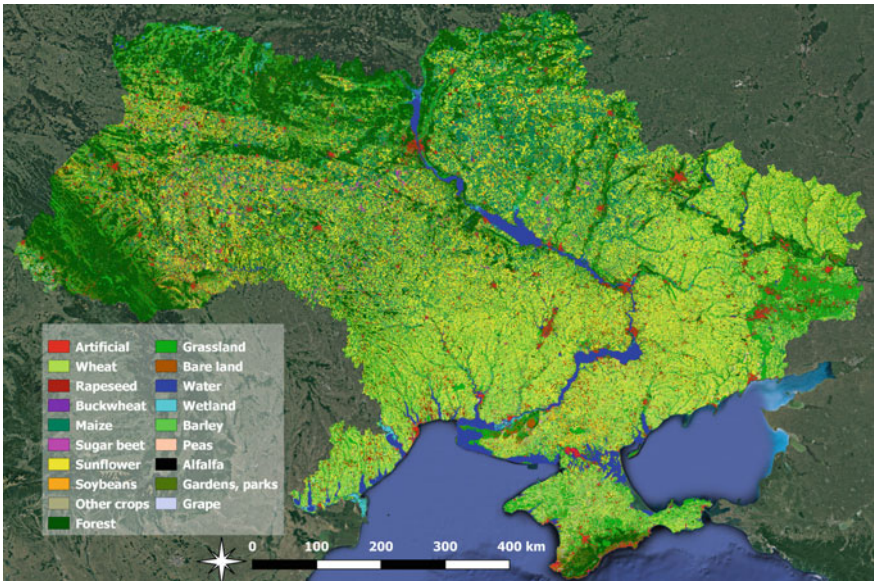



**Fig. 4** Classification map for Ukraine, 2021

| Kappa = 0.94 | Artificial | Cereal crops | Rapeseed | Buckwheat | Maize | Sugar beet | Sunflower | Soy | Other crops | Forest | Grassland | Bare land | Water | Wetland | Peas | Alfalfa | Gardens, parks | Grape | UA | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Artificial** | 22567 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 16 | 0 | 42 | 12561 | 4053 | 39 | 0 | 0 | 26 | 220 | 57,1 | 70,6 |
| **Cereal crops** | 294 | 1865708 | 1355 | 3795 | 3548 | 0 | 1614 | 98 | 5182 | 0 | 10318 | 219 | 263 | 36 | 26334 | 5729 | 7926 | 4422 | 96,3 | 97,8 |
| **Rapeseed** | 0 | 0 | 464897 | 5330 | 0 | 0 | 0 | 0 | 114 | 0 | 0 | 0 | 0 | 0 | 2734 | 0 | 0 | 630 | 98,1 | 98,9 |
| **Buckwheat** | 0 | 0 | 0 | 1823 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 24 |
| **Maize** | 10 | 175 | 0 | 163 | 562293 | 1546 | 2469 | 14863 | 1008 | 0 | 647 | 9 | 17 | 516 | 173 | 20 | 2641 | 1259 | 95,7 | 96,9 |
| **Sugar beet** | 0 | 1 | 0 | 0 | 136 | 32510 | 270 | 1326 | 645 | 0 | 0 | 0 | 0 | 9 | 3 | 0 | 1 | 0 | 93,1 | 86,6 |
| **Sunflower** | 18 | 412 | 0 | 762 | 712 | 4301 | 792949 | 10824 | 2246 | 0 | 2 | 76 | 0 | 691 | 13 | 340 | 322 | 97 | 97,4 | 98,2 |
| **Soy** | 0 | 2662 | 0 | 741 | 4726 | 1732 | 2947 | 143695 | 1441 | 0 | 158 | 4 | 0 | 175 | 138 | 1 | 989 | 5 | 90,1 | 86,9 |
| **Other crops** | 170 | 397 | 88 | 3 | 82 | 78 | 310 | 14 | 4796 | 0 | 80 | 374 | 3550 | 18 | 56 | 0 | 676 | 390 | 43,3 | 34,4 |
| **Forest** | 439 | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 9 | 87527 | 43 | 0 | 13 | 427 | 0 | 0 | 16261 | 0 | 83,6 | 91 |
| **Grassland** | 234 | 8558 | 45 | 536 | 1233 | 0 | 450 | 167 | 580 | 23 | 163312 | 130 | 312 | 6380 | 984 | 11268 | 29861 | 4756 | 71,4 | 80,5 |
| **Bare land** | 404 | 52 | 0 | 0 | 36 | 0 | 0 | 7 | 1 | 0 | 56 | 16558 | 4359 | 4 | 12 | 0 | 268 | 3 | 76,1 | 64 |
| **Water** | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 773905 | 114 | 0 | 0 | 0 | 0 | 100 | 99,2 |
| **Wetland** | 50 | 4 | 0 | 0 | 135 | 0 | 0 | 23 | 12 | 0 | 411 | 3 | 41 | 8682 | 0 | 0 | 39 | 4 | 92,3 | 63,5 |
| **Peas** | 0 | 1 | 31 | 0 | 0 | 0 | 0 | 0 | 472 | 0 | 0 | 0 | 0 | 0 | 33884 | 0 | 0 | 0 | 98,5 | 68,6 |
| **Alfalfa** | 0 | 531 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 9 | 12 | 19662 | 0 | 0 | 97,2 | 68,5 |
| **Gardens, parks** | 166 | 54 | 0 | 197 | 94 | 0 | 184 | 104 | 160 | 192 | 1746 | 34 | 14 | 820 | 0 | 132 | 54795 | 13261 | 76,2 | 59 |
| **Grape** | 32 | 1 | 0 | 4 | 0 | 0 | 10 | 0 | 123 | 0 | 3 | 15 | 5 | 0 | 0 | 0 | 144 | 11059 | 97 | 46,6 |
| **PA** | 92,5 | 99,3 | 99,7 | 13,7 | 98,1 | 80,9 | 99 | 84 | 28,5 | 99,8 | 92,4 | 55,2 | 98,4 | 48,4 | 52,7 | 52,9 | 48,1 | 30,6 | **94,8** | |

**Fig. 5** Confusion matrix for major classes based on test in-situ dataset for RF method

| Kappa = 0.96 | Artificial | Cereal crops | Rapeseed | Buckwheat | Maize | Sugar beet | Sunflower | Soy | Other crops | Forest | Grassland | Bare land | Water | Wetland | Peas | Alfalfa | Gardens, parks | Grape | UA | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Artificial** | 21763 | 139 | 4 | 0 | 0 | 0 | 28 | 5 | 55 | 187 | 653 | 3860 | 2183 | 54 | 0 | 5 | 963 | 46 | 72,7 | 80,1 |
| **Cereal crops** | 5 | 1854789 | 1063 | 12 | 1096 | 0 | 5 | 26 | 228 | 0 | 2744 | 3 | 21 | 331 | 1818 | 920 | 659 | 4 | 99,5 | 99,1 |
| **Rapeseed** | 1 | 1139 | 464795 | 0 | 15 | 0 | 0 | 0 | 491 | 0 | 474 | 0 | 16 | 0 | 226 | 6 | 2 | 0 | 99,5 | 99,6 |
| **Buckwheat** | 2 | 167 | 22 | 13084 | 702 | 20 | 617 | 3451 | 127 | 0 | 192 | 0 | 0 | 39 | 604 | 0 | 0 | 0 | 68,8 | 80,8 |
| **Maize** | 2 | 2 | 0 | 0 | 558356 | 92 | 2646 | 2683 | 290 | 7 | 100 | 1 | 331 | 796 | 0 | 144 | 38 | 0 | 98,7 | 98,1 |
| **Sugar beet** | 8 | 0 | 0 | 124 | 1109 | 39840 | 3848 | 311 | 68 | 0 | 610 | 0 | 9 | 238 | 0 | 0 | 54 | 71 | 86,1 | 92,2 |
| **Sunflower** | 10 | 728 | 0 | 4 | 3764 | 62 | 789533 | 2145 | 630 | 0 | 334 | 169 | 218 | 154 | 74 | 0 | 79 | 7 | 99 | 98,8 |
| **Soy** | 0 | 227 | 0 | 0 | 6636 | 134 | 673 | 160965 | 171 | 0 | 365 | 0 | 1 | 122 | 0 | 13 | 163 | 0 | 95 | 94,5 |
| **Other crops** | 2418 | 8118 | 372 | 51 | 370 | 19 | 1707 | 708 | 12118 | 0 | 1030 | 6610 | 1491 | 493 | 93 | 1807 | 377 | 166 | 31,9 | 44,3 |
| **Forest** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 213 | 86488 | 65 | 0 | 3 | 355 | 0 | 0 | 2718 | 0 | 96,3 | 97,4 |
| **Grassland** | 71 | 3588 | 99 | 9 | 83 | 0 | 459 | 24 | 717 | 0 | 128295 | 9 | 191 | 3845 | 0 | 558 | 8409 | 169 | 87,6 | 79,4 |
| **Bare land** | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 80 | 19269 | 10894 | 0 | 0 | 0 | 0 | 0 | 63,5 | 63,9 |
| **Water** | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 62 | 771007 | 17 | 0 | 0 | 0 | 0 | 100 | 99 |
| **Wetland** | 0 | 0 | 0 | 0 | 111 | 0 | 6 | 1 | 0 | 0 | 1152 | 0 | 106 | 10546 | 0 | 0 | 0 | 0 | 88,5 | 70,8 |
| **Peas** | 0 | 1802 | 4 | 41 | 56 | 0 | 100 | 354 | 891 | 0 | 32 | 0 | 0 | 0 | 61510 | 1094 | 5 | 0 | 93,4 | 94,5 |
| **Alfalfa** | 0 | 1610 | 1 | 0 | 77 | 0 | 0 | 24 | 56 | 0 | 1451 | 0 | 11 | 1 | 0 | 32135 | 136 | 0 | 90,5 | 88,4 |
| **Gardens, parks** | 10 | 6162 | 52 | 25 | 695 | 0 | 1035 | 346 | 639 | 1071 | 38255 | 0 | 35 | 891 | 1 | 394 | 98746 | 1666 | 65,8 | 74,8 |
| **Grape** | 0 | 1 | 0 | 0 | 4 | 7 | 454 | 7 | 88 | 0 | 952 | 2 | 3 | 2 | 0 | 87 | 1537 | 33974 | 91,5 | 92,8 |
| **PA** | 89,2 | 98,7 | 99,7 | 98 | 97,4 | 99,2 | 98,6 | 94,1 | 72,1 | 98,6 | 72,6 | 64,3 | 98 | 59 | 95,6 | 86,5 | 86,7 | 94,1 | **96,7** | |

**Fig. 6** Confusion matrix for major classes based on test in-situ dataset for NN method

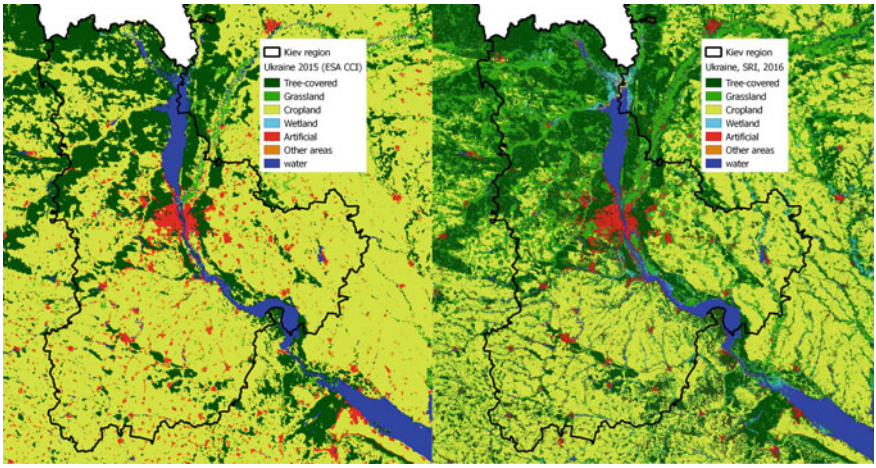



**Fig. 7** Comparison of ESA CCI land cover (left) with land cover (right) over Kyiv region

instance, for Kyiv region of Ukraine the overall accuracy of classification map is improved by more than 10% compared to ESA's Climate Change Initiative Land Cover dataset; the kappa coefficient for ESA's Climate Change Initiative Land Cover dataset is 0.75, while the kappa coefficient for our map is 0.9. Comparison of ESA CCI land cover with land cover over Kyiv region on the base of our NN approach is shown on Fig. 7. Concerning ESA CCI land cover, we have identified significant overestimation of cropland areas (Fig. 8). Using the collected ground truth data along the roads in 2021 the LC/LU map for the territory of Kyiv region was created in Google Cloud Platform (Fig. 9).

The overall accuracy of such a map is 95%, which is for 2% higher than the accuracy of the Random Forest map (Table 4). The differences between such maps are especially noticeable on minority crops. In the future, we plan to use Google Cloud buckets with satellite data and create LC/LU maps for the entire territory of Ukraine using deep learning methods. In the future, the created map of the land cover will solve the problems of transparent land use, implementation of agricultural policy and sustainable development of society in Ukraine.

Using land cover maps for 2016 and 2019 with 10 m spatial resolution we calculated built-up area change for Ukrainian cities and with use of NASA Gridded Population of the World product we obtain this indicator on the city and country scale (Fig. 9). To conduct SDG 11.3.1 calculation for Ukraine, we created the Functional Urban Area layer compliant to the Copernicus Urban Atlas products [34].

Indicator 15.1.1 for 2019–2021 for the regions of Ukraine and for Ukraine as a whole was assessed using the land cover map (Fig. 10). The obtained results are validated with the available information according to the state statistics of Ukraine provided by State statistical Service in Ukraine. The average deviation of the calculated values from the statistical indicators is about 5%.

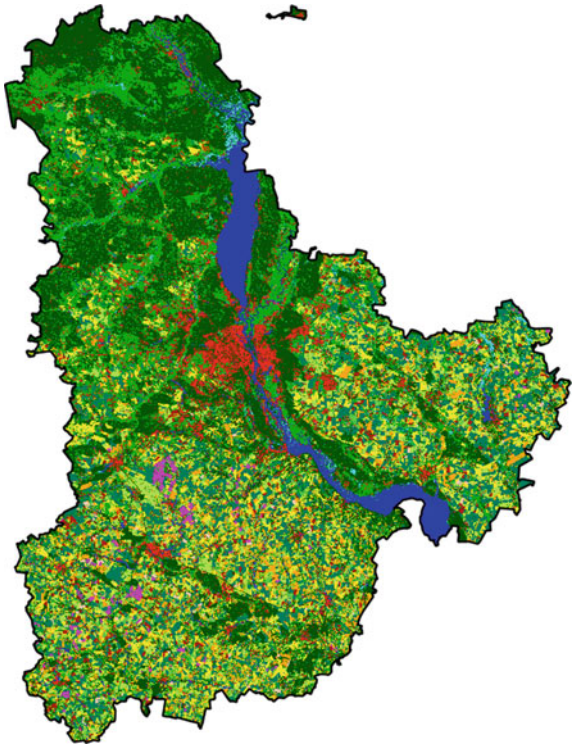


**Fig. 8** The result of classification for the Kyiv region in Google Cloud Platform

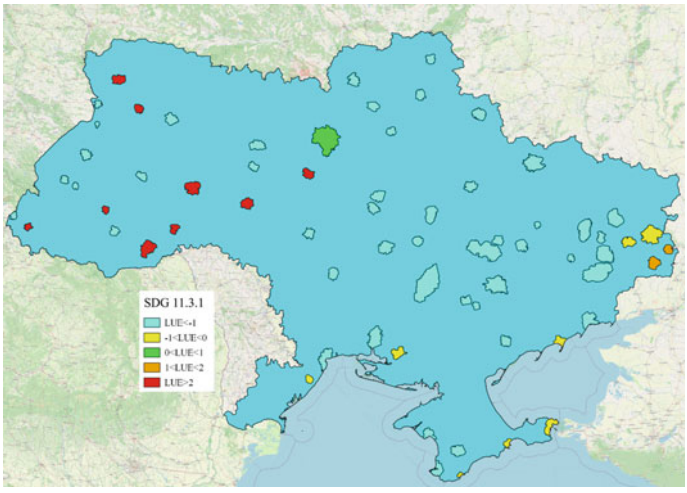



**Fig. 9** The indicator 11.3.1 «Ratio of land consumption rate to population growth rate» for Ukraine

**Table 4** Accuracies comparison of NN and RF algorithms in Google Cloud Platform and Google Earth Engine

| | Google cloud platform (NN) | | | GEE (RF) | | |
|---|---|---|---|---|---|---|
| | UA | PA | **F1** | UA | PA | **F1** |
| Artificial | **92,6** | **85,5** | **88,9** | 82,3 | 81,1 | 81,7 |
| Wheat | **90,1** | **96,9** | **93,4** | 84,4 | 98,3 | 90,8 |
| Barley | **86,9** | **49,3** | **62,9** | 47,6 | 7,4 | 12,7 |
| Rape | 100 | 100 | 100 | 99,1 | 99,8 | 99,4 |
| Maize | 97,5 | 97,4 | 97,4 | 97,1 | 95,3 | 96,2 |
| Sugar beet | 99,6 | 99,9 | 99,8 | 99,5 | 98,7 | 99,1 |
| Sunflower | **96,9** | **98,7** | **97,8** | 96,3 | 95,4 | 95,9 |
| Soy | **94,2** | **89,6** | **91,9** | 87,6 | 90,2 | 88,9 |
| Other crops | **55,8** | **64,4** | **59,8** | 10,6 | 8,5 | 9,5 |
| Forest | 99,8 | 99,7 | 99,8 | 98,8 | 99,8 | 99,3 |
| Grassland | **86** | **97,7** | **91,5** | 78,1 | 94,7 | 85,6 |
| Bare land | 71,6 | 70,9 | 71,2 | 74,5 | 72 | 73,3 |
| Water | 99,9 | 100 | 100 | 100 | 100 | 100 |
| Wetland | 76,5 | 89,9 | 82,7 | 95,7 | 30,6 | 46,3 |
| Peas | **100** | **89** | **94,2** | 100 | 31,6 | 48 |
| Overall accuracy | **95,8** | | | **93,1** | | |

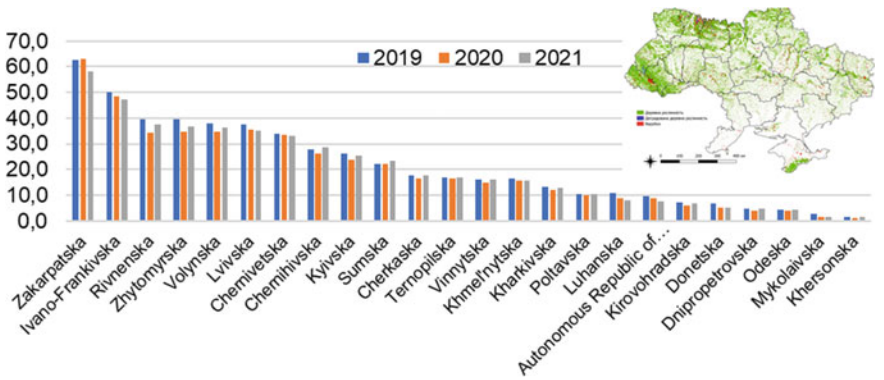



**Fig. 10** The indicator 15.1.1 «Forest area as proportion of total land area» for Ukraine

And, finally, as another valuable product, which is based on classification maps, the preliminary land degradation map for whole Ukraine with spatial resolution 30 m was created based. For that we have used the trend analysis of vegetation indices of Landsat-8 and Sentinel-2 satellite data. This product is the basis for the SDG 2.4.1 and 15.3.1 indicators calculation and in this process, its creation is the most time-consuming task (Fig. 11).

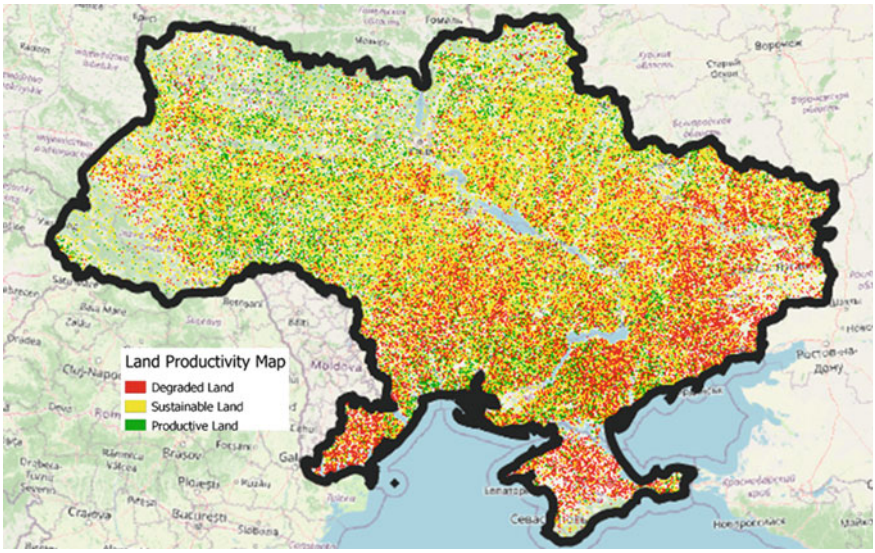



**Fig. 11** The land productivity map for Ukraine

## 5 Conclusions

To summarize aforementioned, it should be stress one more time, that for all the geospatial products we have used modern information technologies and resources, namely, cloud infrastructures (Amazon Web Services, Google Earth Engine etc.) with different kinds of resources such as Elastic Computing 2 (EC2), Simple Storage Service (S3) and others. These processing tools were provided within joint pilot projects of GEO Committee and key market players, in particular Google and Amazon companies. For programming and AI models development we realized Python scripts within Anaconda and IPython Notebook software on the base of standard packages and specialized ones' like rasterio, gdal, boto3 and others, which are very useful for model realization, data uploading, preprocessing, accuracy estimation etc.

All the geospatial products for SDG indicators estimation are based on a number of basic geospatial products, namely classification maps for different years [29], DEM [30], vector OSM data. Besides satellite data, which are accessible within cloud infra-structures and ready for usage, we have collected in-situ data and prepared in-situ datasets for models training and validation [31]. For classification map overall accuracy was about 95%, where for all the classes accuracy was from 75 to 98%. On the base of obtained results, we can generate country-level geospatial products, which are more accurate than on the base of global or regional dataset with coarse resolution. In particular, on the base of our maps it can be created different products for SDG indicators estimation, degradation maps [32], productivity maps [33], air quality estimation products [34] and solved another important applied tasks.

On the base of classification maps [35] we can build additional products, for instance, to estimate the levels of land degradation with spatial resolution 10 m [36]. Given that it is proposed to use data with a spatial resolution of 10 m to obtain geospatial products, the question of time series optimization based on, in particular, estimation methods described in [37] will also be investigated in the future.

## References


1. Belgiu M, Ovidiu C (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens Environ 204:509–523
2. Menno-Jan K, Ricker B, Engelhardt Y (2018) Challenges of mapping sustainable development goals indicators data. ISPRS Int J Geo Inf 7(12):482
3. Compendium of Earth Observation contributions to the SDG Targets and Indicators. https://eo4society.esa.int/wp-content/uploads/2021/01/EO_Compendium-for-SDGs.pdf. Accessed 18 Feb 2022
4. Open Data Cube (ODC). https://www.opendatacube.org/. Accessed 18 Feb 2022
5. Diepen CA, Rappoldt C, Wolf J, Keulen H (1988) Crop growth simulation model WOFOST. Documentation version 4.1. for World Food Studies. Wageningen, The Netherlands
6. Food and agriculture organization of the United Nations. Indicator 2.4.1 - Proportion of agricultural area under productive and sustainable agriculture. https://www.fao.org/sustainable-development-goals/indicators/241/en/. Accessed 18 Feb 2022
7. Indicator 11.3.1. https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf. Accessed 18 Feb 2022
8. Indicator 15.1.1. http://www.fao.org/sustainable-development-goals/indicators/1511/en/. Accessed 18 Feb 2022
9. Indicator 15.3.1. https://knowledge.unccd.int/topics/sustainable-development-goals-sdgs/sdg-indicator-1531. Accessed 18 Feb 2022
10. Kussul N, Lavreniuk M, Shumilo L (2020) Deep recurrent neural network for crop classification task based on Sentinel-1 and Sentinel-2 imagery. In: IGARSS 2020 IEEE international geoscience and remote sensing symposium, Waikoloa, HI, USA, pp 6914–6917. https://doi.org/10.1109/IGARSS39084.2020.9324699
11. Copernicus Sentinel mission. http://copernicus.eu/main/sentinels. Accessed 18 Feb 2022
12. State Statistics Service of Ukraine, http://www.ukrstat.gov.ua/, last accessed 2022/02/18
13. JECAM Guidelines for cropland and crop type definition and field data collection. http://jecam.org/wp-content/uploads/2018/10/JECAM_Guidelines_for_Field_Data_Collection_v1_0.pdf. Accessed 18 Feb 2022
14. Shelestov A, Lavreniuk M, Vasiliev V, Shumilo L, Kolotii A, Yailymov B, Kussul N, Yailymova H (2020) Cloud approach to automated crop classification using sentinel-1 imagery. IEEE Trans Big Data 6(3):572–582. https://doi.org/10.1109/TBDATA.2019.2940237
15. Shelestov A, Lavreniuk M, Kussul N, Novikov A, Skakun S (2017) Exploring google earth engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. Front Earth Sci 5:17
16. Shelestov A, Lavreniuk M, Kolotii A, Vasiliev V, Shumilo L, Kussul N (2017) Cloud approach to automated crop classification using Sentinel-1 imagery. In: Proceedings of the 2017 conference on Big Data from Space (BiDS 2017), pp 122–125
17. Sims NC, et al (2020) A land degradation interpretation matrix for reporting on UN SDG indicator 15.3.1 and land degradation neutrality. Environ Sci Policy 114:1–6
18. Kussul N, Lavreniuk M, Kolotii A, Skakun S, Rakoid O, Shumilo L (2020) A workflow for Sustainable Development Goals indicators assessment based on high-resolution satellite data. Int J Digital Earth 2(13):309–321. https://doi.org/10.1080/17538947.2019.1610807

19. Kaifeng P et al (2021) Evaluating the potential impacts of land use changes on ecosystem service value under multiple scenarios in support of SDG reporting: a case study of the wuhan urban agglomeration. J Clean Prod 307:127321
20. Rizvi SR, et al (2020) Data cube application algorithms for the united nation sustainable development goals (UN-SDGS). In IGARSS 2020 IEEE international geoscience and remote sensing symposium. IEEE
21. Hamilton SE, Friess D (2018) Global carbon stocks and potential emissions due to mangrove deforestation from 2000 to 2012. Nat Clim Chang 8(3):240–244
22. Kussul N, et al (2019) Nexus approach for calculating SDG indicator 2.4.1 using remote sensing and biophysical modeling. In IGARSS 2019 IEEE international geoscience and remote sensing symposium. IEEE
23. Vignola GB, Acs S, Borchardt S (2020) Modelling for sustainable development goals (SDGs): overview of JRC models
24. Kussul N, Shelestov A, Yailymova H, Yailymov B, Lavreniuk M, Ilyashenko M (2020) Satellite agricultural monitoring in Ukraine at country level: world bank project. In: IGARSS 2020–2020 IEEE international geoscience and remote sensing symposium, pp 1050–1053
25. Kussul N, Shelestov A, Lavreniuk M, Yailymov B, Kolotii A, Yailymova H, Skakun S, Shumilo L, Bilokonska Y (2020) SDG indicator 11.3.1 within HORIZON-2020 SMURBS, Space research in Ukraine, 2018–2020, pp 91–95
26. Shelestov A, et al (2020) Assessment of land consumption for sdg indicator 11.3. 1 using global and local built-up area maps. In IGARSS 2020–2020 IEEE international geoscience and remote sensing symposium
27. Global Human Settlement layer. https://ghsl.jrc.ec.europa.eu/datasets.php. Accessed 18 Feb 2022
28. Forest resources assessment. http://www.fao.org/forest-resources-assessment/en/. Accessed 18 Feb 2022
29. Shumilo L et al (2021) Is soil bonitet an adequate indicator for agricultural land appraisal in Ukraine. Sustainability 13(21):12096
30. Liu X (2008) Airborne LiDAR for DEM generation: some critical issues. Prog Phys Geogr 32(1):31–49
31. Kussul N et al (2012) Crop state and area estimation in Ukraine based on remote and in-situ observations. Int J Inf Models Anal 1(3):251–259
32. Kussul N, et al (2017) Land degradation estimation from global and national satellite based datasets within UN program. In 2017 9th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS), vol 1
33. Kussul N, et al (2017) Cropland productivity assessment for Ukraine based on time series of optical satellite images. In 2017 IEEE international geoscience and remote sensing symposium (IGARSS)
34. Shelestov A et al (2021) Air quality estimation in ukraine using SDG 11.6.2 indicator assessment. Remote Sens 13(23):4769
35. Kussul N, Shelestov A, Lavreniuk M, Butko I, Skakun S (2016) Deep learning approach for large scale land cover mapping based on remote sensing data fusion. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS), pp 198–201
36. Kussul N, Lavreniuk N, Shelestov A, Yailymov B, Butko I (2016) Land cover changes analysis based on deep machine learning technique. J Autom Inf Sci 48:42–54
37. Azarskov VN, Blohin LN, Zhitetskii LS, Kussul NN (2004) Robust methods for estimation, identification and adaptive control. NAU, p 500. (in Russian)

# The Comprehencive Approach to Big Data Preprocessing




**Larysa Globa, Rina Novogrudska, and Mariya Grebinichenko**

**Abstract** Nowadays, Big Data research is making significant progress. The paper is devoted to optimizing the process of Big Data pre-processing. The existing shortcomings of input datasets that lead to a decrease in their quality in systems that Big Data processing have been identified. The main methods of pre-processing of data sets are considered. The ways to Big Data clearing are described using of which allows to correct distorted data. The existing approaches ways to designing the architecture of Big Data processing systems are analyzed and microservice architecture was used for their flexible processing. The possibilities of Big Data pre-processing have been expanded due to the improved method of data clearing based on the text data processing templates. The proposed advanced flexible complex of algorithms for Big Data pre-processing with a high level of fault tolerance allows increasing the accuracy of data further processing. Software realization (web-applications) of proposed algorithms complex for data cleansing methods with proposed improvements and microservice architecture was developed. The efficiency of the proposed architecture for the Big Data pre-processing system based on microservices is shown on practice.

**Keywords** Big Data · Preprocessing · Data cleaning · Algorithm · Text data

L. Globa (✉) · R. Novogrudska (✉) · M. Grebinichenko (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: lgloba@its.kpi.ua

R. Novogrudska
e-mail: rinan@ukr.net

M. Grebinichenko
e-mail: grebinichenko.maria@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_6

## 1 Introduction

Nowadays, there is no single definition of Big Data. At this time, it is not possible to determine exactly how much data can be considered as Big. However, the existence of Big Data itself is not beneficial: further impact on the data and the subject domain with which it is related, have only the result in analysis or their use in machine learning. Such Big Data research is making significant progress and even could change human lives.

In 2021, IBS company estimated that the total amount of data in the world is about 40 zettabytes. Internet users alone generate about 2.5 quintillion data per day (Fig. 1) [1].

Undoubtedly, the term "Big Data" means a large amount of information, but in addition to obtaining and establishing the existence of a wide range of data, an integral part was a set of approaches, mathematical methods, software and hardware that can perform its further processing [2].

In technical systems, incorrect data can enter the system due to a number of different factors, such as temperature effects, mechanical or software failure, conditions that are not provided for the correct operation of sensors, overloading of the processing system, etc.

As a result, invalid, erroneous, omitted values may occur in the final datasets, making it virtually impossible to obtain high-quality Big Data calculations.

The purpose of proposed research is to increase the efficiency of Big Data processing through the use of an advanced method of Big Data preprocessing, which eliminates invalid values.

The paper is structured as following: Sect. 2 gives main characteristic features of Big Data, analyses of related works and backgrounds for the research, as well as approaches to Big Data quality improvement. Section 3 describes modified algorithm of data preprocessing. In Sect. 4, main features and steps of data cleaning method

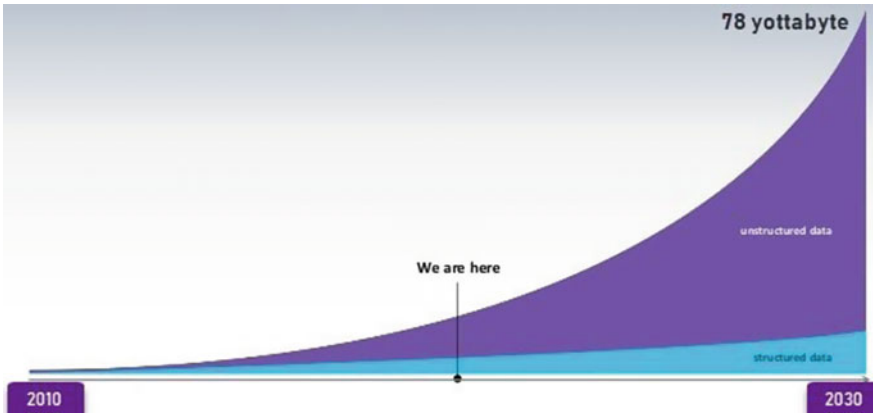



**Fig. 1** The growth of information over time

proposed by authors are given. Section 5 presents prototype of data cleaning system, including its architecture and components description. In Sect. 6 conclusions and plans for future work are given.

## 2 State of Art and Backgrounds

### *2.1 Characteristic Features of Big Data*

The importance of Big Data does not depend only on the actual amount of data, but mainly on how it is used. Taking data from any source and analyzing them, you can solve the following problems [3]:

1. simplified resource management;
2. increasing of operations efficiency;
3. optimization of product development;
4. creating new opportunities for profit and growth of the company.

Combining Big Data with high-performance analytics, you can find solutions to the following examples of business problems:

- Identify the root causes of failures, problems and defects in near real time.
- Rapid response to anomalies.
- Improving patient outcomes through rapid conversion of medical imaging data.
- Improving the ability of deep learning models to accurately classify variables and respond to them.
- Detect fraudulent behavior before it has consequences.

Nowadays different methods of artifisial intelligence is also used for Big data storage, analisis and processing. Such methods includes: machine learning [4, 5], data minning [6, 7], ontological modeling [8, 9], statistics [10, 11], etc.

Before using Big Data, it is in need to think about how the process of working with Big Data in the network of systems and users will unfold. There are five key elements that together can get the most out of a Big Data factory [3]:

1. Big Data strategy.
2. Identify the Big Data source.
3. Big Data access, manage and store.
4. Data analysis.
5. Qualitative conclusions based on data.

However, all the above advantages of working with Big Data are effective and appropriate only in the context of working with high quality data that do not contain false values, and therefore cannot make a significant error in the final results and influence the wrong conclusion.

## 2.2 Data Preprocessing

The set of techniques used prior to the application of a data mining method is named as data preprocessing for data mining [12] and it is known to be one of the most meaningful issues within the famous Knowledge Discovery from Data process [13]. Since data will likely be imperfect, containing inconsistencies and redundancies is not directly applicable for a starting a data mining process. It is in need to mention the fast growing of data generation rates and their size in business, industrial, academic and science applications. The bigger amounts of data collected require more sophisticated mechanisms to analyze it. Data preprocessing is able to adapt the data to the requirements posed by each data mining algorithm, enabling to process data that would be unfeasible otherwise [14].

In [15] one of the first books on preprocessing in Big Data a large amount of significant issues are described, namely the enumeration and description of some of the most recent solutions to address imbalanced classification, the characteristics of novel problems and applications (with the latest published algorithms), and the implementations of working techniques ready to be used in well-known Big Data.

The data is to be transformed in the form of understandable and useable insights by algorithms and models. The data mining steps require data that is cleaned and structured to a larger extent. This is achieved by using various algorithms, processes and applications known as data pre-processing techniques. [16] reviews various data pre-processing techniques from a big data point of view.

Researches in [17] studied a process for big data analysis, and proposed an efficient methodology of entire process from collecting big data to implying the result of big data analysis. In addition, patent documents have the characteristics of big data, thus in the paper an approach to apply big data analysis to patent data is proposed, and the result of patent big data to build R&D strategy is implied. As well a case study is performed using applied and registered patent documents retrieved from the patent databases in the world. This allows illustrating how to use proposed methodology for real problem.

## 2.3 Approaches to Improving the Big Data Quality

[18] analyzed a number of the most effective and common mathematical methods and tools that can be used for preprocessing of Big Data. Since this study aims to improve the work with textual data, let us present the main aspects of data given text format processing [19].

Detection and removal of particles and meaningless words. Text formations that do not have their own meaning and/or words that are ignored by search engines only create additional workload and resource consumption, but do not affect the final result of the analysis. Software removal of service parts of speech is performed by

existing libraries of neurolinguistic data analysis. The most common of them are NLTK (Python) and phpMorphy (PHP, porting to Node.JS).

Remove punctuation from text. Modern text recognition and word processing systems, unfortunately, cannot recognize the meaning of punctuation marks, so last are ignored to reduce the impact of additional information noise on the data. This operation can be performed by defining arrays of valid characters (Cyrillic, Latin, numeric characters if necessary), and step-by-step analysis of data and the content of illegal characters. When entering too large data sets, character-by-character analysis can take an extraordinary amount of time, so page-by-page analysis is used: the dataset is broken into relatively small parts and processed in parallel in different streams. In this case, additional intermediate datasets should be created, and after all threads are completed, the initial data should be replaced by the obtained calculation results.

Convert text to lowercase. In this step, the initial text data is formatted to a single format, and then makes it impossible to create duplicate copies of the same words. This calculation is done by encoding ASCII text characters. The lowercase case occupies a certain interval in the notation of bit sequences (01,100,001–01,111,010). When deleting all invalid non-text data, you must check that the required text character is within this range, and change the fifth bit of the code to the reverse character. This will change the code value by 32 bits to match the same lowercase character.

Lemmatization. Stemmatization. Lemmatization means the process of transforming words into their canonical form: the form of the infinitive, the nominative case, the singular. Natural text processing algorithms are currently unable to recognize shades of words expressed by different cases, and words in different cases do not change their meaning. In addition, the semantic meaning of words is also ignored during preprocessing. Stemmatization is a related method that aims to remove word-building elements that have almost no effect on the basic meaning of words, but significantly reduce their total number and, consequently, the load during further processing [20].

The analysis of all mentioned above aspects showed that their use in the algorithm for datasets cleaning makes it possible to improve the overall quality of the original data. However, this brings only a partial improvement to the overall algorithm, it is necessary to study the system as a whole to improve it at the architectural level.

## 3 Modified Algorithm of Data Preprocessing

In a previous study [18], existing methods and approaches for data cleaning were analyzed. All the most common preprocessing methods are essentially aimed at correcting only one aspect of contaminated data. The main idea of the modified data set cleaning algorithm proposed in the study is to add an additional step to the main process, the purpose of which is to combine different methods into a single processing algorithm.

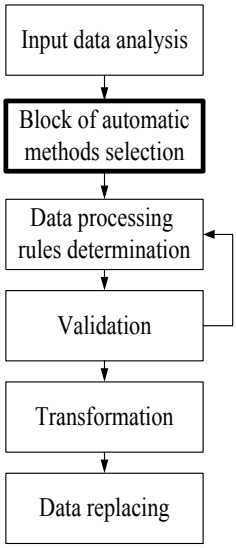



**Fig. 2** Block diagram of the data cleaning process

However, not all data can be collected from different sources, and therefore contain different model structures and have different relationships between parameters, or contain missing data, because the data collection sensors worked high quality throughout the period of operation [18]. Otherwise, there may be no duplicates in the dataset, because the database from which the set was compiled does not contain duplicate records, or they could be deleted manually by an expert. Therefore, the existing methods of preprocessing of Big Data are not a universal solution—for each set of data to be processed in each case, the appropriate optimal cleaning process must be established.

The cleaning process consists of the following stages (Fig. 2) [21]:

1. Analysis of the provided data: at this stage a set of metadata about the provided arrays of information is created manually or with the help of software implementation.
2. Automatic selection of data cleaning methods.
3. Defining the rules and procedures of data processing: determining the method to be applied to a given set of information, checking the availability of all necessary related data necessary for its implementation (for example, parameters set by the expert).
4. Validation: data is checked or cleaned according to the generated metadata, whether the processing is correct and efficient; performed iteratively with the analysis (often—on a certain part of the data) to achieve the best result.
5. Transformation: after confirming that the data cleaning is correct and effective, the changes are applied to the entire data set.
6. Replace data: to update the data in the source, where the data will later be accepted for processing, cleared correct data replaces them in the repository.

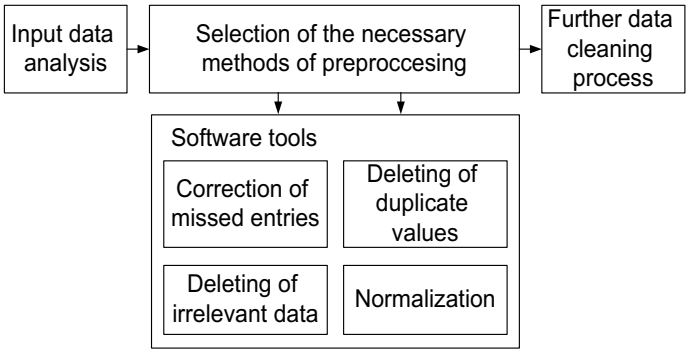



**Fig. 3** This caption has one line so it is centered

The proposed modified data cleaning algorithm addresses the following aspects of data contamination (Fig. 3):

1. Recovery of damaged and missed records.
2. Delete duplicate entries.
3. Correction of peak and irrelevant values.

The proposed modified algorithm aims to analyze the input data and adaptively create a unique procedure for cleaning. Clear selection criteria must be established for the selection of algorithms for processing. The efficiency of the modified algorithm was proved in comparison with standard methods of cleaning large volumes of data [18], but it had a number of certain significant shortcomings:

1. Work only with numerical data.
2. Formation of the cleaning algorithm each time, regardless of whether the same data has already been processed.

Proposed modified data cleaning algorithm allows correcting the above shortcomings.

## 4 Advanced Data Cleaning Method

Support for working with text data is undoubtedly able to significantly expand the functionality and capabilities for data analysis. Most datasets today are not just about numeric values.

However, maintaining and processing text datasets requires large computational resources and the implementation of machine learning to clear semantic meaning. There are hundreds of different principles and methods of working with data, some of which were discussed in the previous section, and even the introduction of most of them in the application will not be able to cover all theoretically possible versions of

text values in the dataset. In addition, existing software components will not be able to work with text values and code duplication will be required. Converting numeric values to text will also not make sense because mathematical methods are used to determine the validity of a data set.

Therefore, the best solution is to reduce and transform text values to numerical ones. Methods of normalization of text data and their transformation into numerical ones will be used for this decision. We need to identify certain limitations that will apply to this approach:

- the value of each indicator will have a fixed number of values;
- the semantic meaning of the data will not be taken into account.

The reduction will consist two stages. The first stage is preparation, normalization of values to a sigle type and format with help of a number of procedures. And the second is the transformation.

The first stage includes the following steps:

1. Marking and deletion of partials and words without sense. Words that don't have their own meaning and (or) words that are ignored by search systems. For example, for English words it may be "a", "an", "the", for Ukrainian—"ti", "ale", "under" and others. In average, approximately 130 words are ignored in English.
2. Punctuation removing. Any punctuation marks must be removed to reduce the general noise of the dataset. If the dataset is exported in.csv format, the presence of division characters can disrupt the general structure of the file, and then the data won't be valid. Removing punctuation marks doesn't affect cleaning and hardening analysis because the semantic structure of the data isn't taken into.
3. Converting text to the lower register. This option formats the information to a single view, and reduces the likelihood of creating repeated copies of the same words. Due to this cleaning procedure, such an occurrence isn't possible.
4. Lemmatization. This point is related to the previous one. By lemmatization we mean the transformation of words to their canonical view: the form of the infinitive, the intervening in line, one. Nowadays text processing algorithms can't recognize word variations expressed by different inditions, diminutive suffixes, persons, etc.
5. Stemmatization. Spore to the previous item, which performs the process of finding the basis of the word for given inputs. This technique is intended to significantly reduce the total number of text elements as well as the amount of workload for further processing.

In this way, the textual values of the indicators will be maximally simplified, which will be easy to compare one to another.

The next step will be the actual reduction of the received values to text values. Since we have introduced the limitation on the number of values, we can now present the textual indicator in the format of the converted type. In this case, each type of dataset is related to one and only one value of the specified indicator. According to this condition, the following steps of converting textual values to numerical ones will be carried out:

1. Determination of the number of uniqueness values of a renegotiated type.
2. Reconfiguration of the dataset format: instead of the text value, N new indicators of the Boolean type will be added, i.e. will have values 0 or 1.
3. Determine the value to which the given tuple belongs by the value of one, and all other newly created indicators by 1. Thus, each tuple has one and only one unit value for the given indicator.

Let's look at how each of the existing components works with the new data format:

1. Filtration of data: on the one hand, text data may not have peaks, minimal or preliminarily mistaken values. However, metadata can contain a list of values that can take a given indicator. Or this list can be set additionally by the expert. In this case the data filtration will take place at the text data normalization stage, because it will not be possible to match the tuple with a certain value. Moreover, it is possible to filter the data if the valued of the indicator for this tuple couldn't be normalized—a high probability that the data is corrupted or was entered incorrectly. So, the data filtration doesn't make sense if the text values are included, which will help to optimize the dynamic algorithm.
2. Renewal of missing values: records with NULL values belong to the records that require cleaning, but they always have a specific value due to metadata definitions. Records that all entries contain NULL don't have to be allowed to purge, because this case if highly likely a complete loss of information. In addition, such records don't have a general impact on the final result of big data processing. When you work with text, it is possible to update values according to the data keys' parameters, or according to the value specified by the expert.
3. Removing doublers: this component has two basis rules. First: if the duplicates are forbidden by some parameter, they will be searched for and deleted. Also, this applies to existing duplicate records or duplicate objects in general. Another thing: duplicates will be rechecked for the primary key (if the metadata or expert has it). First we check the data for the specified indicators in the key, and then we check other records without them, i.e. we make sure that the records will not contain fully identical data if we can get the key parameters.

So, all existing components are logically and technically combined by reducing textual values to numerical ones. Let us analyze the possible effectiveness of the suggested modification to form the conditions for including this component in the modified purification algorithm.

Changing the structure of the initial dataset and adding new values and columns of indicators, it is important to keep the approximate size of the initial set of data or change it within a certain standard tolerable deviation. We will try to estimate such deviation, in which further data processing will remain optimal.

The general formula for the dependence of changes in the size of the dataset can be defined as:

$$K = \left( \sum_{i}^{T} x_i \right) * n$$

where K is the final size of parameters in the changed dataset, T—number of parameters with text value in the initial dataset, x—number of values for the i-th indicator, n—number of records in the dataset.

Let us first look the dependence of the number of indicators for a single column. Here a linear dependence between the number of indicators and the newly created columns takes place.

$$K_0 = a * n$$

where $K_0$ is the size of parameters in the initial dataset, a—average length of the value of this parameter, n—number of records in the dataset.

After creating new columns, they take values of only zero or one, i.e. the value can be equated to the atomic value. Thus, after the columns are reduced, we obtain the following size of the parameter:

$$K = x_i * n$$

It is possible to create N−1 columns instead of N. To determine whether the tuple corresponds to one of its values, we define at most one indicator as 1, all other new columns have the value 0. If the tuple corresponds to 0 in all indicators, it means that it corresponds to the N-th value of the parameter.

$$K = (x_i - 1) * n$$

As the size of the own announcement of the new column depends on the format of data representation, we will take it out with this indicator. Thus, the size of a single text indicator will be changed by a / (xi−1). Basing on this fact, we can formulate the following statements:

1. If a $\geq$ xi, the size of the dataset will be reduced, which reduces the time and resource consumption of further processing and is an acceptable modification.
2. If a < xi, the size of the dataset will increase as well as the resources that need to be spent on data cleaning and further processing. Such conversion may not be optimal and must be prohibited by the order, or be performed only after verification and confirmation of the conversion by an expert.

Let us consider the general case when more than one indicator can take a text value. Let us express the initial size of the text data through the average length of the value of each individual indicator:

$$K_0 = \left( \sum_{i}^{T} a_i \right) * n$$

where K—initial size of parameters in the changed dataset, T—number of parameters with text value in the initial dataset, a—average length of values for the i-th indicator, n—number of records in the dataset.

Change the formula according to the last transformations:

$$K = \left( \sum_{i}^{T} (x_i - 1) \right) * n$$

where K is the final size of parameters in the changed dataset, T—number of parameters with text value in the initial dataset, x—number of values for the i-th indicator, n—number of records in the dataset.

The size of the text values in the final data set will be changed to:

$$\frac{K}{K_0} = \frac{\left( \sum_{i}^{T} a_i \right)}{\left( \sum_{i}^{T} (x_i - 1) \right)}$$

We define this ratio as N0. This value reflects the threshold value for converting the dataset according to the order. So, the algorithm can be included in the execution under one of two conditions:

1. If the threshold parameter Z0 is not set by the expert, the algorithm will be executed only under the condition of $0 \leq \frac{K}{K_0} \leq N_0$.
2. If the threshold parameter Z0 was set by the expert, the algorithm will be executed only under the condition of $0 \leq \frac{K}{K_0} \leq Z_0$.

Advantages of the formulated cleansing component:

- Support for textual data when cleaning data sets;
- Use of existing components for working with text values with minimal changes;
- Data set optimization (working with logical values instead of textual ones, data set size reduction).

Disadvantages of the created approach:

- There are certain restrictions on the processing of text data (a large number of unique values of the text parameter).

After performing such steps of the modified algorithm as “Conversion” and “Data substitution”, it is possible to convert the dataset into the previously formed structure. This is useful when the next data processing should work with text values. When working on the interrelationships between dataset entries, this step is irrelevant or even unnecessary, since additional resources will be used. That’s why this step can be included or excluded by the expert (for the request of the imposed one).

This step will take place because the parameters of the converted dataset have names in the format “Indicator_name”, so after the data cleansing it will be possible

to form the headers of the group by the initial name of the indicator and write to the final column the value of the adjacency of the tuple to a certain value.

Thus, we get an additional program component for preliminary data cleansing. Below is a table listing all the preprocessing software components, their working principles and conditions for their inclusion in the process (Table 1).

**Table 1** Components for dynamic preliminary data processing

| Cleaning component | Data type | Principle of work | Automatic startup options |
|---|---|---|---|
| Data filtration | Numeric | Removing irrelevant item values. After creating the scheme of the data object, the user is asked to enter the maximum and minimum values for each of them | The number of records that aren't relevant under initial conditions is greater than $Z_0$ |
| Renewing of missing values | Numeric | Filling the missing parameter values in the data sets. The changes are filled with the average value for the parameter | The number of missing records that can't get NULL value greater than $Z_0$ when the threshold value is set by the expert |
| Deletion of doubles | Numeric, string | Identification of repeated records in the dataset | • If the duplicates aren't allowed by a certain parameter, they will be searched for and deleted. Also the presence of duplicate objects in general, or repeated entries<br>• Checking duplicates for primary key (if metadata has one). First the data for the specified fields in the key are checked, and then—all records without them, that is, if we get the parameters of the key, or there will not be fully identical data in the records |
| Reduction text values | String | Text fields can take three classes A, B and C, for each row in the dataset one of three given columns (A, B, C). On output we get 3 columns that have appropriate names (A, B, C), and for each row in the dataset zeros and one in the column to which it belongs | When test values are converted, the dimension of the dataset changes no more than N0 timer per order ($Z_0$ when set by the expert) |

Most of Big Data systems have type—specific or irregular data patterns. Typically, experts analyze data that was collected from the same data sources, but for different circumstances or at different time intervals. The expert wants to carry out a comparative analysis of the work of a certain system in different sections, and therefore the sets of data have the same or similar structure.

Let us consider as an example an analysis of the data obtained during the operation of certain server machines. This set of data contains the following indicators:

- Sample number—any natural number.
- Energy to be used by the server.
- The frequency of data processing.
- The number of streams.

The expert can collect data about the operation of server machines at different loads, during different calculations, under certain temperature conditions or data about the operation of different server machines. But the indicators for the research will remain the same after the data processing. In this case, there is no sense to conduct the automatic selection of cleaning tools each time the data is loaded.

To increase the total preprocessing time, it is proposed to generate input dataset metadata at the input stream analysis stage and store it on the server [22]. After forming the preprocessing algorithm in the step "Defining processing rules", the rules in the declarative view are added to the previously generated metadata and the record on the server is updated (Fig. 4).

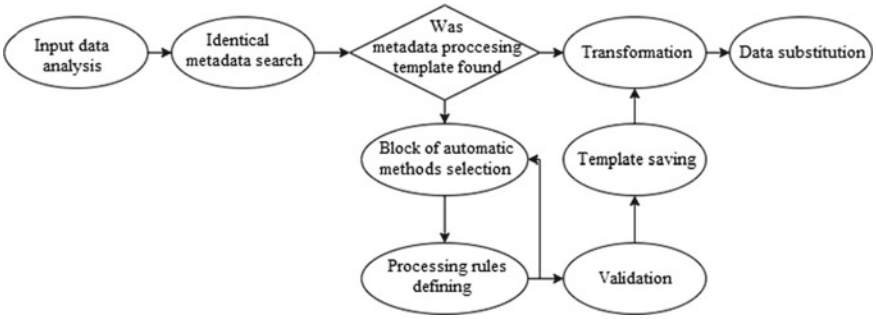



**Fig. 4** Block diagram of the modified data cleaning process

```
{
    metadata: {general_dataset_info},
    data_schema:[
        property1: {...},
        property2: {...},
        ...
    ],
    cleanse_schema:[
        algorithm1: {[algorithm_parameters]},
        algorithm1: {[algorithm_parameters]},
        ...
    ]
}
```

During the next request to the server, metadata of the new flow should be compared with already stored metadata. If you find a match, the "Definition of processing rules" block is skipped, and cleaning is performed according to the already formed rules. Thus, while storing a relatively small amount of data, this approach gives a significant gain in processing time by increasing the level of automation.

The advantages of the requested method are:

1. Increasing the reliability of the final result of the work.
2. Improvement of quality of input data through loss correction and removal of wrong records.
3. Reducing the time for \data pre-processing through to use of processing templates.

The disadvantaged of the suggested method are:

1. Using additional resources for pre-processing data.
2. Using additional server memory to store processing templates.

## 5 Data Cleaning System Prototype

### *5.1 Data Cleaning System Architecture*

Before starting to develop a new functional, it is necessary to find the necessary architecture that best meets the requirements of the add–on and, if the architecture changes, perform refactoring of the existing functional, test it and only then add new components.

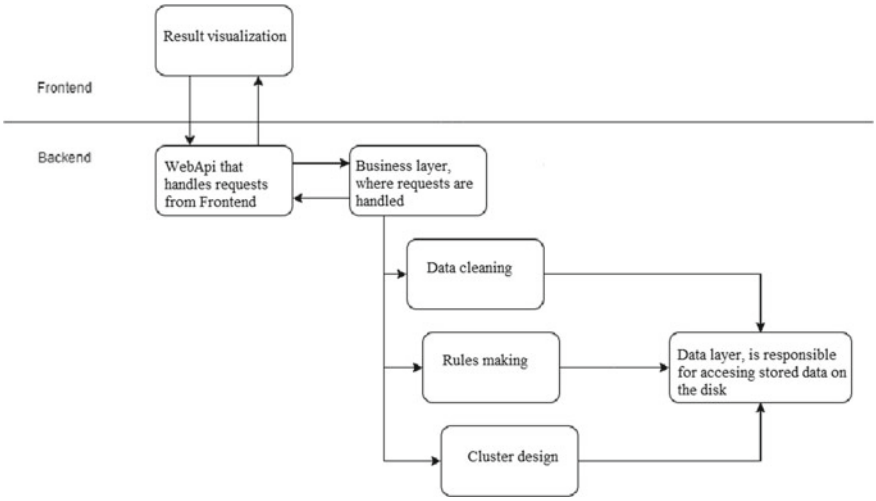



**Fig. 5** Monolithic architecture of the first version of the application

Figure 5 shows the architecture of the first prototype add—on for data clustering. The implemented program has a classical monolithic design with one database, settled on two servers—for server and client part.

The following requirements were formed for the developed application:

1. Each of the main analytical components (data cleaning, rules making, cluster design) must work independently of each other.
2. Each of the listed components must store intermittent calculation results and return them to users on demand.
3. Each of the listed components can provide access to its functionality via external API (e.g., for further extension of the functionality, adding a new interface, etc.).

This client implementation is primarily designed to form and visualize clusters for clients. In this case, microservices architecture is costly and only generates additional costs. However, the components have a strong link to one other. The second and third conditions can be done by combining the three components into a single API.

While researching framework for writing the client–side part of the add–on, it was decided to rewrite the Angular part with React to increase the speed of the program. In addition, it was decided to keep two client add-ons: for visualization of clusters (Angular 10) and visualization of non—white rules with the possibility of transformation into mobile add—on (React). Thus, the architecture should look as follows Fig. 6.

At once we can distinguish two main disadvantages of such implementation. First of all, due to the fact that the server part now receives twice as many requests, the probability of rejection also increases. Moreover, if the server crashes, two of the client add—on will not be able to continue to work. It is impossible not to mention the database: despite the fact that all records to the database are burned in the transaction,

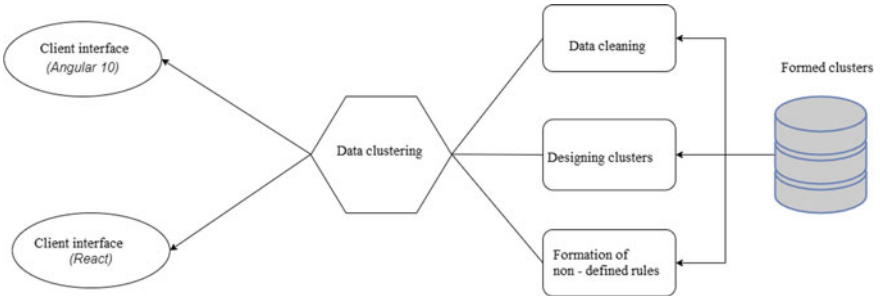



**Fig. 6** A new service—oriented architecture

there is a low probability of data corruption (e.g., power failure, physical damage to the server, ets.). In such cases, data will be lost for all components, both cleaning and clustering as well as forming imprecise rules.

To solve these problems, we suggest using micro services architecture (Fig. 7). By dividing the server part into separate add-ons, the flexibility and load on each individual server will be improved. In additional, it will be possible to use different technology stack for each component in order to turn the most efficient tools to solve the problem.

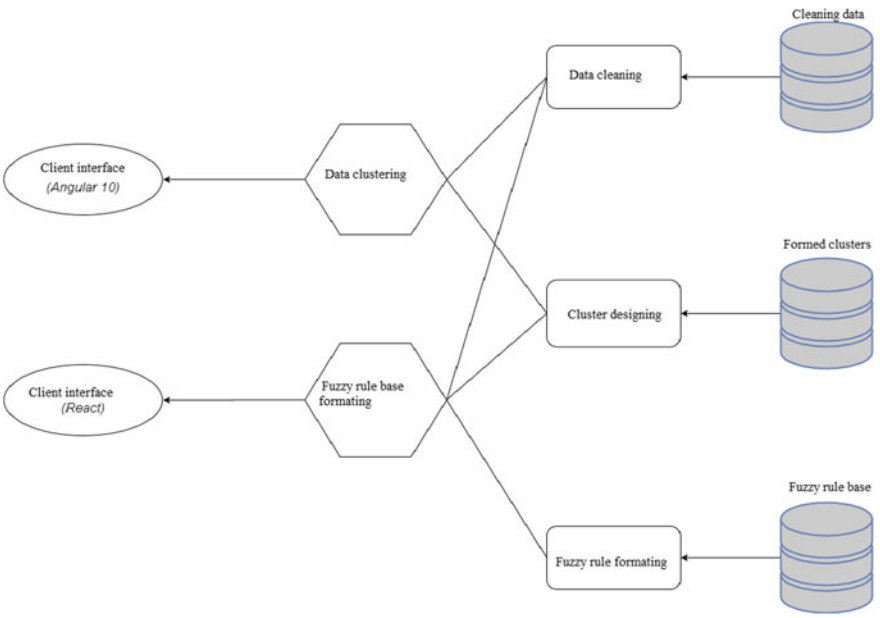



**Fig. 7** End-to-end architecture with microservices

Let's divide the monolith into three main components:

1. Data cleaning—server addendum in.Net Core and Python, storing cleaning templates, processed data (if necessary).
2. Design of clusters—server add—on.Net Core, storage of processed data, formed clusters.
3. Formation of partial rules—server add—in Node.js, storage of partial rules.

A separate database will be developed for each component, each add—on will function on a separate server, and communication will be set up using HTTP requests. In addition, only the components that will be used will be configured for each add–on.

When breaking the code base into micro servers, the main functionality will be transferred withour problems, because each component is independent and doesn't have strong links one from another. However, a component with service functions (e.g., preparing an input file for processing, exporting a file, service calculations, etc.) is required for each program component. It is not possible to develop it on a separate server, because it is not a stand—alone software component with a specific purpose. Moreover, to work with files (e.g., formatting editing values), the function must be located on the same server, because otherwise the data must also be sent for formatting.

Therefore, this feature makes the code on each of the denoted servers be doubled. This causes additional consumption of resources, but due to a small number of service functions, this duplication is acceptable and doesn't affect the overall consumption of resources.

To investigate and compare the above mentioned architecture, three variants of the add—on were deployed in the Azure environment. The functionality of this service allows you to temporarily dispose of add—ons on the servers, due to which the comparison was made. In order to determine the conditions of server degradation, the capacity and power of the servers was reduced.

In order to obtain the results when modeling this situation, the data processing note to each individual component was performed: data cleaning (D), clustering (C) and formation of fuzzy rules (F). Situations such as error return (−), lack of response from the server (−), perceptible response delay (±), normal operation (+), minor response delay (+) were monitored.

Let's look at the result of modeling of service interruptions (Table 2).

Based on these results, we can conclude that the monolithic architecture has an extremely low level of fault tolerance. Due to the fact that all components are deployed within one solution on one server, all other components couldn't work when one of them is disconnected, which prevents the first condition of independent operation of each component from being fulfilled.

The other servo—oriented architecture showed significantly better results in the study of interoperability, since each service is developed independently from each other and doesn't have strong links in the software implementation. As expected, duplication of the service code didn't affect the overall algorithms performance of any of the components. However, we can distinguish two main disadvantages associated with executing transactions on the database. When we reduced the capacity of one of

**Table 2** Comparison of the architecture of the developed addendum

| The situation that modeled | Is it possible to operate components in such conditions? | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Monolithic architecture | | | New server—oriented architecture | | | Microservices architecture | | |
| | D | C | F | D | C | F | D | C | F |
| Invalid DB | − | − | − | − | − | − | ± | ± | ± |
| Invalid service D | − | − | − | − | + | + | − | + | + |
| Invalid service C | − | − | − | + | − | + | + | − | + |
| Invalid service F | − | − | − | + | + | − | + | + | − |
| Overload D | ± | − | − | ± | ± | ± | ± | + | + |
| Overload C | − | ± | − | ± | ± | ± | + | ± | + |
| Overload F | − | − | ± | ± | ± | ± | + | + | ± |

the component servers and increased the number of requests to it, we saw a significant increase in processing and response time. This was due to long execution of requests in transactions to the database. In addition, when the database is broken, the extension couldn't perform one of the specified conditions, namely saving intermediate results, so the result of this situation was categorized as negative in the correlation table.

Finally, the third algorithm showed the best results. By storing the data in separate storages when the database is cancelled, the information of other components will still be available, so the addendum will remain partially operable. In addition, when one of the components is reloaded, you can write to other components as quickly as transactions are made to different data sources.

Therefore, the micro services architecture was chosen for refactoring and further implementation of the functional as the most efficient one for accomplishing the tasks set.

## *5.2 Program Component for Text Data Transformation*

After selecting the architecture, refactoring and testing the already implemented functions, it is necessary to develop a new component for working with text data. The main change to the already created implementation of the modified algorithm is the introduction of the function of work with text values of the indicators. The development process was started by analyzing already existing text processing tolls for delegating tasks of initial processing and text normalization to third—party libraries.

Choosing the following variants for text processing:

1. LemmaGenerator –.NET Core, no longer supported. Easy to use, but the list of functions is very limited and there is no documentation on how to work with the library.
2. NLTK library, or NLTK—a package of programs and libraried for symbolic and statistical processing of natural language, written in Python programming language.
3. Lemma Morphological Analyzer—Lemmatic Analyzer written in Elixir programming language, language based on Erlang. It is implemented using the classical recorder method which is based on an abstraction called Finite State Transducer. It has a wide set of features, but it is not recommended for production use, because this library isn't effective for both the CPU and memory.
4. PhpMorphy—morphological analysis library, which is implemented in PHP programming language. PHP isn't supported by partitioned operations. In most cases, a load balancer is used to maintain the load across multiple servers at both program and data levels. Split processing is a necessary condition for increasing the amount of data to be processed.
5. Node Phpmorphy—full functional porting phpMorphy on Node.JS. Contains all the advantages of the framework without the disadvantages of the PHP platform, but is neoptimized library because the porting method is only the transfer of the functionality without taking into account the intricacies of the platform.

Therefore, due to low benefits and a small number of drawbacks, the NLTK library was chosen. Let's look in detail at the microservices that are responsible for the front—end data processing (Fig. 8).

This component contains one microservice (block of modified algorithm,NET Core) and one additional service (transformation of text data, Python). Another element isn't a microservice because it doesn't have a separate logically and clearly defined business problem, which could fulfill. In addition, it doesn't have its own database, because it evaluates the input stream and returns the intermediate values of data prepared for cleaning. That is why the service was logically added to the microservice of data cleaning.

We conducted an experiment to determine the effect of improved processing of text data on the dataset. The investigation was carried out using the manual adjustment of the algorithm, for the qualitative assessment of the changes all methods of data cleaning were included. The experiment was conducted in two stages: first, the data were cleaned and clustered without text data conversion (Fig. 9), and after that—with conversion (Fig. 10). The number of clusters was equal to eight. After the first step, a graphical representation of the created clusters was obtained. We can conclude from the diagrams (Fig. 11) that the clusters were formed clearly, without any markers, but they weren't formed quite clearly.

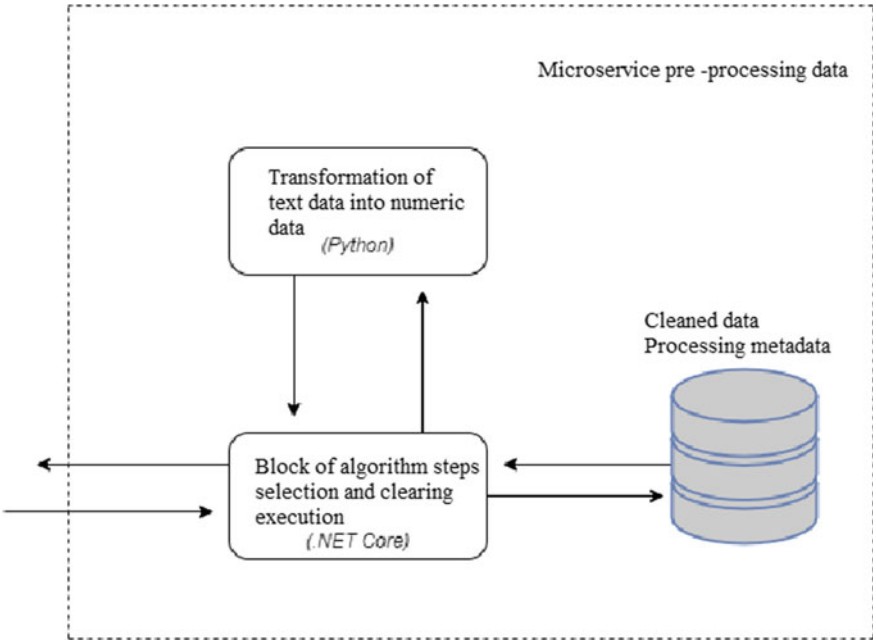



**Fig. 8** Microservices for pre-processing data

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Examples | FR | TR | EN | OS | |
| 2 | 215 | 1300 | 2 | 637.727 | Windows | |
| 3 | 1335 | 2400 | 4 | 3448.939 | UNIX | |
| 4 | 1756 | 2800 | 1 | 522.3634 | Windows | |
| 5 | 754 | 1800 | 1 | 3755.148 | UNIX | |
| 6 | 1357 | 2400 | 2 | 329.2791 | NetWare | |
| 7 | 1644 | 2800 | 16 | 362.1201 | UNIX | |
| 8 | 1261 | 2400 | 16 | 87.9331 | Windows | |
| 9 | 59 | 1200 | 32 | 457.0115 | NetWare | |
| 10 | 1892 | 2901 | 16 | 622.3048 | NetWare | |
| 11 | 1304 | 2400 | 32 | 643.8911 | UNIX | |
| 12 | 1033 | 2200 | 8 | 661.5897 | Windows | |
| 13 | 1270 | 2400 | 16 | 76.89376 | UNIX | |
| 14 | 958 | 2000 | 4 | 291.2624 | UNIX | |

**Fig. 9** Data structure before processing

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Examples | FR | TR | EN | OS_UNIX | OS_NetWare |
| 2 | 215 | 1300 | 2 | 637.727 | 0 | 0 |
| 3 | 1335 | 2400 | 4 | 3448.939 | 1 | 0 |
| 4 | 1756 | 2800 | 1 | 522.3634 | 0 | 0 |
| 5 | 754 | 1800 | 1 | 3755.148 | 1 | 0 |
| 6 | 1357 | 2400 | 2 | 329.2791 | 0 | 1 |
| 7 | 1644 | 2800 | 16 | 362.1201 | 1 | 0 |
| 8 | 1261 | 2400 | 16 | 87.9331 | 0 | 0 |
| 9 | 59 | 1200 | 32 | 457.0115 | 0 | 1 |
| 10 | 1892 | 2901 | 16 | 622.3048 | 0 | 1 |
| 11 | 1304 | 2400 | 32 | 643.8911 | 1 | 0 |
| 12 | 1033 | 2200 | 8 | 661.5897 | 0 | 0 |
| 13 | 1270 | 2400 | 16 | 76.89376 | 1 | 0 |
| 14 | 958 | 2000 | 4 | 291.2624 | 1 | 0 |

**Fig. 10** Data structure after processing

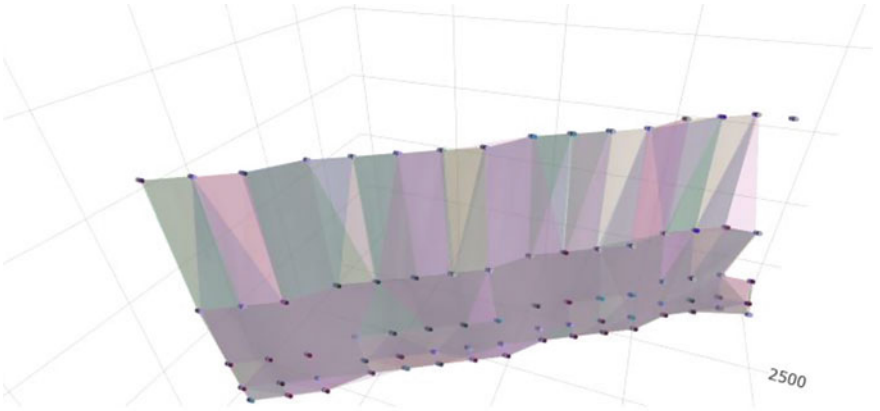



**Fig. 11** The result of clustering the data before the transformation

On the second stage, the diagram of the formed clusters looks clearer (Fig. 12). Not only the data transformation, but also the fact that other components of data cleaning influenced the search for and correction of invalid data.

We compare the number of computing iterations that the clustering algorithm required to form the final processing result. During the research we collected the following data on Fig. 13.

The implementation of the reduced processing and text transformation functionality helps to improve the quality of the output data and the final result of big data processing, as well as to reduce the number of iterations of clustering, and, consequently, the consumption of software and numerical data.

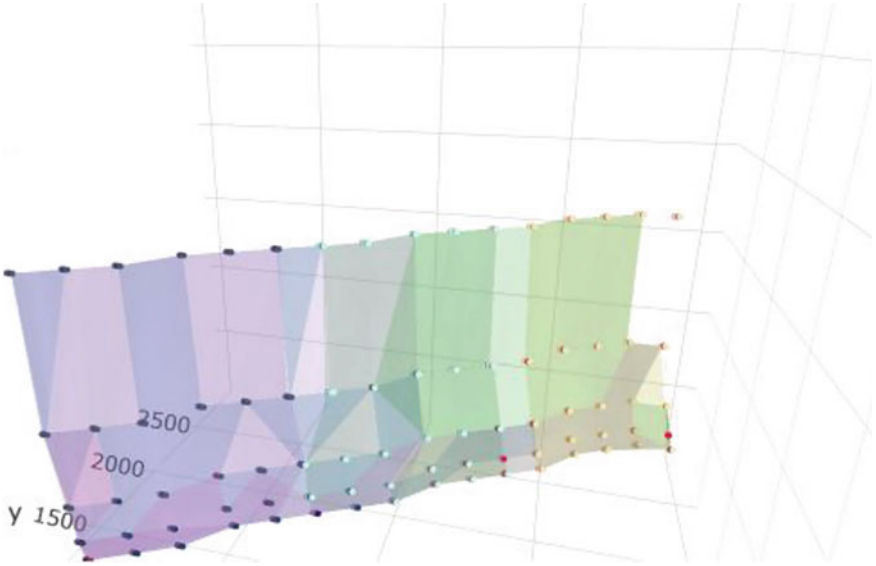



**Fig. 12** The result of clustering the data after the transformation

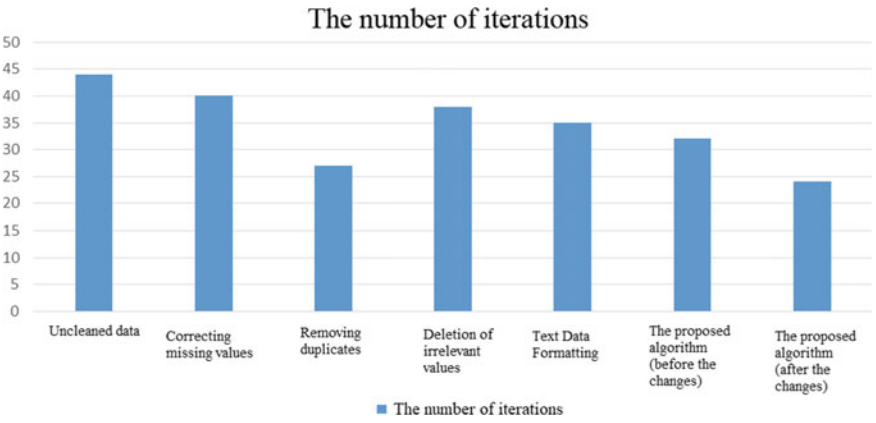



**Fig. 13** A comparative diagram of the number of iterations during clustering

## *5.3 Implementation of Data Cleaning Templates*

We decided to implement data cleaning templates in the existing component of the dynamic algorithm, because we need access both to the final data and crops of the algorithm, as well as an open connection to the database. The template will be stored in a separate table in JSON format.

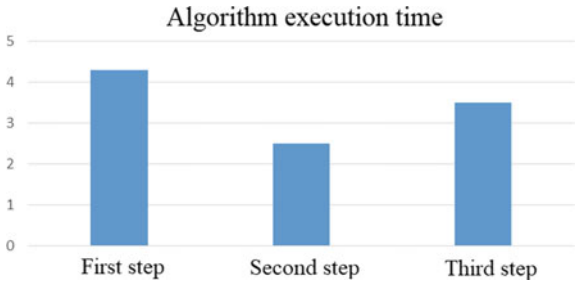



**Fig. 14** Comparative diagram of the time to execute the data cleaning algorithm

Before processing begins, the expert can choose whether the processing template will be used, or whether the sequence of crops for clearing should be reconstructed. The comparative analysis was performed in three stages:

1. Loading the dataset, selecting the cleaning terms, forming the template.
2. Loading the dataset, using a template.
3. Loading the dataset, ignoring the cleaning template.

As a result of the study, we measured the approximate time of work execution for the data pre-processing blocks and obtained the following results on Fig. 14.

Thus, the use of data processing templates allows to significantly reduce the time for data pre-processing, as well as to avoid additional use of accounting resources.

## 6 Conclusions

The paper proposes an improved algorithm of data preprocessing, which allows improving the process of its preprocessing by dynamically determining the software component that corresponds to the type of a particular dataset. The advanced method of text data clearing is based on its reduction to a format in which preprocessing by existing components of the preprocessing system is possible. Also, the mechanism of saving metadata of the dataset is developed.

During the research the layout of the Big Data cleaning system was created, with specific architecture including the component for text data transformation, which allow processing large amounts of information and as a result increase the reliability of the results of analytical Big Data processing due to the proposed modified algorithm.

Future researches will be aimed at subsequent testing of Big Data cleaning system with large amount of distributed data of different formats and types to identify system problems and their elimination.

## References


1. Big Data. https://www.it.ua/knowledge-base/technology-innovation/big-data-bolshie-dannye. Accessed 03 Jan 2022
2. What is Big Data? Introduction, Types, Characteristics, Examples. https://www.guru99.com/what-is-big-data.html. Accessed 27 Dec 2021
3. Big Data. What it is and why it matters. https://www.sas.com/en_ca/insights/big-data/what-is-big-data.html
4. Wu Q, Ding G, et al (2016) A survey of machine learning for big data processing. Eurasip J Adv Signal Process 67. doi:https://doi.org/10.1186/s13634-016-0355-x
5. Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: opportunities and challenges. Neurocomputing 237:350–361. https://doi.org/10.1016/j.neucom.2017.01.026
6. Wu X, Zhu X, Wu G-Q, Ding W (2014) Data mining with big data. IEEE Trans Knowl Data Eng 26(1):97–107. https://doi.org/10.1109/tkde.2013.109
7. Fan W, Bifet A (2013) Mining big data: current status, and forecast to the future. Sigkdd Explor Newsl 14:1–5. https://doi.org/10.1145/2481244.2481246
8. Popova M, Novogrudska R (2021) Cognitive load of ontology as a means of information representation in the educational process. In: Arai K (eds) Intelligent computing. Lecture notes in networks and systems, vol 285. Springer, Cham. https://doi.org/10.1007/978-3-030-80129-8_25
9. Globa L, Gvozdetska N, Novogrudska R (2021) Ontological model for data processing organization in information and communication networks. Syst Res Inf Technol, 47–60. https://doi.org/10.20535/srit.2308-8893.2021.1.04
10. Galeano P, Peña D (2019) Data science, big data and statistics. Test 28:289–329. https://doi.org/10.1007/s11749-019-00651-9
11. Dunson DB (2018) Statistics in the big data era: failures of the machine. Stat Prob Lett 136:4–9. https://doi.org/10.1016/j.spl.2018.02.028
12. García S, Luengo J, Herrera F (2015) Data preprocessing in data mining. Springer, Berlin
13. Han J, Kamber M, Pei J (2011) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann Publishers Inc., Burlington
14. García S, Ramírez-Gallego S, Luengo J et al (2016) Big data preprocessing: methods and prospects. Big Data Anal 1:9. https://doi.org/10.1186/s41044-016-0014-0
15. Luengo J, García-Gil D, Ramírez-Gallego S, García S, Herrera F (2020) big data preprocessing. enabling smart data. Springer Nature Switzerland AG, p 186. https://doi.org/10.1007/978-3-030-39105-8
16. Prakash A, Navya N, Natarajan J (2019) Big data preprocessing for modern world: opportunities and challenges. In: Hemanth J, Fernando X, Lafata P, Baig Z (eds) International conference on intelligent data communication technologies and internet of things (ICICI) 2018. ICICI 2018. Lecture notes on data engineering and communications technologies, vol 26. Springer, Cham. https://doi.org/10.1007/978-3-030-03146-6_37
17. Sunghae J (2015) A big data preprocessing using statistical text mining. J Kor Inst Intell Syst 25(5):470–476
18. Grebinichenko M (2021) Methods of big data preprocessing. Master's thesis, Kyiv, p 54
19. Beginner's Guide to Data Cleaning and Feature Extraction in NLP. Accessed 23 Dec 2021. http://towardsdatascience.com
20. Karimov R, Samkova M, Nikitina S, Akinin A (2016) Using a hybrid algorithm for lemmatization of a diachronic corpus. In: CEUR workshop proceedings, vol 1886, pp 1–8
21. Martin RC (2017) Clean architecture: a craftsman's guide to software structure and design, 1st edn, p 352
22. Grebinichenko M (2021) Data cleansing for increasing performance in IOT networks. In: XV international scientific conference "modern challenges in telecommunications" MCT-2021, conference proceedings. Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, pp 235–237

# Mathematical Models and Informational Technologies of Crop Yield Forecasting in Cloud Environment




**Leonid Shumilo, Sofia Drozd, Nataliia Kussul, Andrii Shelestov, and Sergiy Sylantyev**

**Abstract** The agricultural sector plays a very important role in the country's economy. Often, crop failures are the cause of protracted financial crises, which are very difficult to overcome. Fortunately, modern scientific methods allow us to predict the yield of certain crops in selected fields based on data on soil properties and weather conditions. Such study requires high-resolution data, the downloading and processing of which requires the involvement of powerful cloud infrastructures, graphics processors and other technological solutions. Involvement of these technologies and reliable forecasting of yield help to make the right decision about sowing and avoid crop failure. However, collecting data on soil properties directly from agronomists and local farmers is a very long and hard process. In addition, the data collected may be inaccurate and expensive. As a result, a process will be launched that will eventually lead to an agrarian and financial crisis. However, a solution to this problem has already been found. Remote sensing data together with

L. Shumilo (✉)
University of Maryland, College Park, MD 20742, USA
e-mail: shumilo.leonid@gmail.com

S. Drozd (✉) · N. Kussul (✉) · A. Shelestov (✉)
Institute of Telecommunication Systems, National Technical University of Ukraine, "Igor Sikorsky Kyiv Polytechnic Institute", Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: sofi.drozd.13@gmail.com

N. Kussul
e-mail: nataliia.kussul@gmail.com

A. Shelestov
e-mail: andrii.shelestov@gmail.com

S. Sylantyev (✉)
Space Research Institute National Academy of Sciences of Ukraine and State Space Agency of Ukraine, Glushkov Avenue 40, 4/1, Kyiv 03680, Ukraine
e-mail: sylantyev@nubip.edu.ua

S. Drozd · N. Kussul · A. Shelestov
Institute of Physics and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_7

biophysical models has long been used in the world. In this section the state-of-the-art and current results in this area have been done.

**Keywords** Cloud computing · Biophysical modeling · Satellite data processing · Yield forecasting

## 1 Introduction

The Food and Agriculture Organization (FAO) of the United Nations estimates that 50% more food needs to be produced by 2050 in order to feed the increasing world population. Many of the regional adaptation options connected with crop yield forecasting identified by COP-26 in Glasgow in October 2021 are largely extensions or intensifications of existing risk-management or land production-enhancement activities. For cropping systems, there are many potential ways to alter management to deal with projected climatic and atmospheric changes. One of them, which was widely discussed in COP 26 was the sentinel monitoring programs and modern information technologies in using seasonal climate forecasting to reduce production risk.

Informational technologies are an important component in the field of contemporary agriculture, agrometeorology, and other technological innovations in the area of crop yield forecasting on the Ukrainian market. However, in addition to information technology, the ongoing forecast efforts to increase food production are deeply connected with regional climate changes and different accurate remote sensing data with a combination of in-situ measurements. It is, therefore, crucial to not only accurately forecast crop yield but also to model and characterize by actual satellite data the processes involved by understanding the meteorological and land physics drivers of crop yield volatility. From point of view of a meteorological, soil physical, and vegetation variables influence crop growth, development, and final yield, the crop yield forecasting model is state-of-the-art with the deep nonlinear connection and with complex interaction [1–5]. These variables are accounted for in both agro-climatic satellite data models as well as vegetation indices assessment in crop yield forecasting statistical models to estimate future crop yield [2, 5, 11].

While agro-climatic models require accurate and objective (and not always available from agronomists and local farmers) information recent increase in the availability of global satellite Sentinel-2 and Landsat [5, 6, 13, 16] observations and advancements in statistical methods WOFOST [3, 4], have fueled the application of machine learning (ML) models [7] on the calculation bases of different cloud platform AWS and GEE [14, 15]. In particular, such model architecture for crop yield forecasting may have the new capability of accounting for additional factors reducing growth and yield on the bases of reliable forecasting of yield help to make the right decision about sowing and avoid crop failure [5, 6, 11]. For better calculation and provide more objective information for crop yield forecasting on the wealth of agro-climatic information we propose to apply high-resolution Sentinel-2 data with powerful cloud infrastructures and graphics processors [5, 6, 10, 11]. Here, we

focus on the automatization crop yield forecasting environment with using modern mathematical models, informational technologies on powerful cloud platform.

It is well-known that modern models are very complex due to data used amount and algorithms complexity. That is why powerful cloud solutions are needed for models use. By deployment of the system on the cloud platform [13], we can overcome challenges of big data problem downloading and processing. For instance, Amazon platform provides easy and fast access to Sentinel-1 and Sentinel-2 imagery via scalable Amazon Web Services (AWS) infrastructure (both storage S3 and computational instances EC2) to solve big data downloading and storing problem and also provides powerful computational resources which are necessary for running advanced deep learning and machine learning methods.

## 2 Data Used

At the beginning of our study the primary task is the need to obtain data for processing. In our case satellite images of high resolution, without clouds, shadows, noise and other defects are extremely useful for the correct conduct of the experiment. Thus, we could not limit ourselves to one data source and do without pre-processing of satellite images and used different sources to obtain satellite data and tools for their processing. In our study, we used open data of remote sensing from satellites Sentinel-2 and Landsat.

Sentinel-2 is the space mission of two optical Sentinel-2A and Sentinel-2B satellites of the European Union of the Copernicus program, optical images of which are characterized by high spatial resolution (10–20 and 60 m) and temporal resolution of five days, when the territory of interest is revisited every five days [16]. Another Copernicus satellite Sentinel-1 with synthetic aperture radar (SAR) provides data collection in all weather conditions, time of day/night. Above high latitudes, the Sentinel-2 passes overlap, and some regions are observed twice or more every 5 days. Each satellite Sentinel-2A and Sentinel-2B has a powerful camera that captures the Earth's surface in 13 spectral ranges.

The mission is mainly used for remote observations and monitoring of agriculture, forests, recording changes in the Earth's vegetation, tracking land use patterns and tracking the effects of various natural disasters. The mission provides Sentinel-2 free and open data distribution. In our study, we took data with a spatial resolution of 10 m (Fig. 1).

The main problem in obtaining images from the Sentinel-2 satellite is cloudiness, which due to weather conditions often makes it impossible to use some images (Fig. 2). Therefore, using only one Sentinel-2 mission is not enough.

However, along with Sentinel the only free satellite mission on a regular basis also with a fairly high spatial resolution is a Landsat mission. In our study, we used Landsat-8 images. It rotates around the Earth every 99 min and has a 16-day revisit time. This is the latest launch of the Landsat family satellite, which is equipped with Operational Land Imager (OLI) instruments and a thermal infrared

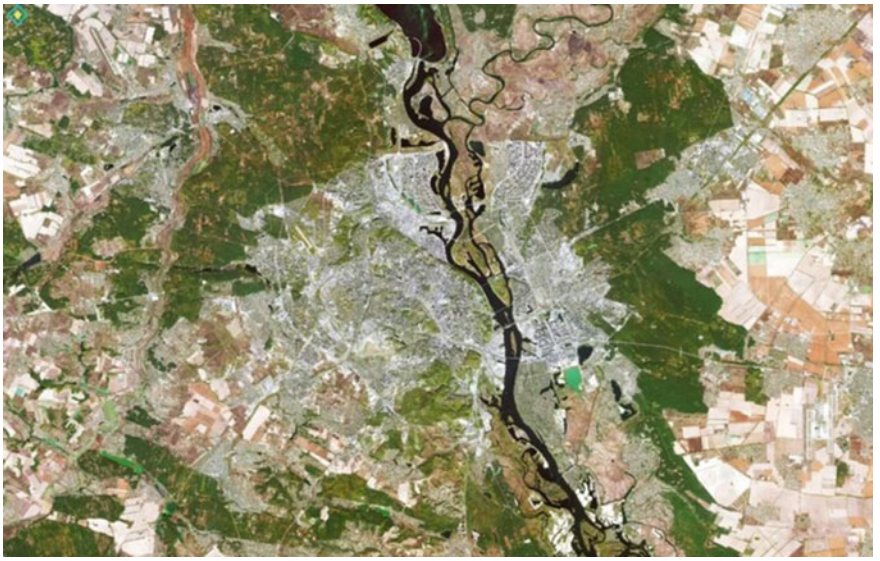


**Fig. 1** Processed satellite image of the Kyiv region with Sentinel-2 for 2021



**Fig. 2** Overcast satellite image of Kyiv region from Landsat-8, 2015

sensor (TIRS). But the Landsat-9 satellite was recently launched and should be available in the near future. These two sensors provide seasonal coverage of global land with a spatial resolution of 30 m (visible, NIR, SWIR); 100 m (thermal channel); and 15 m (panchromatic mode). OLI captures data with improved radiometric accuracy in the 12-bit dynamic range, which improves the overall signal-to-noise ratio. We used satellite images with a pixel size of 30 m.

Data from Sentinel-2 and Landsat satellites can be downloaded using the Amazon S3 repository of the Copernicus Open Access Center. The following pre-processing steps were used for optical data: calibration, stacking, atmospheric correction, peripheral correction, and original raster scaling. To combat the cloudiness of optical data, Landsat-8 (30 m) and Sentinel-2 (10 m) data were merged with joint registration of products from different satellites and scaling of values.

Vegetation indices NDVI and DVI for 2016–2021 were received based on medium-resolution satellite data with high temporal resolution MODIS sensors aboard the Terra and Aqua satellites (Figs. 3 and 4). Data from both satellites are generated at 16-day intervals with different spatial resolution (250–1000 m) [17].

MODIS consists of two spectrometers, one of which (MODIS-N) takes off, and the axis of the other (MODIS-T) can be deflected. 36 MODIS spectral bands cover the range with wavelengths from 0.4 to 14.4 μm. Shooting in two zones is carried out with a resolution of 250 m, in five zones of the visible and near infrared range with a resolution of 500 m, and in the other—1000 m. Data with a spatial resolution of 250 m were used in this research.

The value of the LAI indicator is very important for our study, since it can be retrieved also using CGMS crop yield forecasting model. It characterizes not only

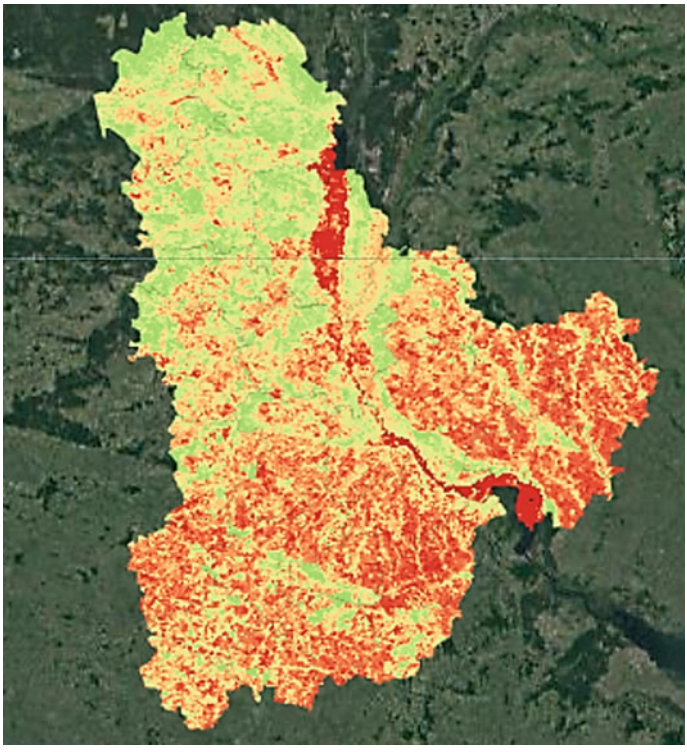


**Fig. 3** Kyiv region MODIS NDVI for 2020



**Fig. 4** Kyiv City MODIS NDVI for 2020

the state of crops related to meteorology, crop type and soil type, but also takes into account (indirectly) land degradation as a factor in reducing land productivity and proper land management as a factor in increasing land productivity.

To evaluate the satellite LAI harmonized data from two satellites were used—MODIS and Landsat-8. The combination of data from both satellites made it possible to build harmonized time series LAI, increasing the temporal resolution and spatial resolution to 30 m.

Estimation of projected without impact of land degradation (ideal case) LAI was conducted using the WOFOST model (CGMS with grid). The model uses crop type parameters, soil type information, meteorological data and standard agro-management rules as input data. At the output for a specific place in the usual national network is obtained LAI for types of crops corresponding to meteorology on soil type. At the intersection of this exit with the map of the type of crops, culture-specific time series are obtained LAI with a high temporary resolution (daily) and a wide resolution of 10 m. The size of CGMS grid cell depends on the weather and the resolution of the soil data.

## 3 Yield Forecasting State-of-the-Art

Remote sensing data are already commonly used for the yield forecasting. With the use of adequate statistical data, this task could be solved even by use of classical machine learning approaches with high accuracy. The most common task is yield forecasting on the rayon or region level. Usually, it is the regression task and the regressors in it are: Vegetation indices such as NDVI, DVI and etc. [1], biophysical parameters such as LAI [2] obtained using remote sensing data, modeled date that reflects crop growth and state [3] and agro-climatic data such as air temperature, land surface temperature and precipitation [4]. Usually, the best way to build high quality yield forecasting model is combination of regressors with different types [5]. But in the case of field level crop yield forecasting, the data requirements are changing significantly. The yield forecasting on the field level has the next main differences:

1. High variation of the data in the small range.
2. Lack of useful historical information for the fields.
3. Small scale of data.
4. Limitation for the number of fields leads to the models overfitting.

The most commonly used data collection for the yield forecasting are based on the moderate resolution satellite data with high temporal resolution, such as MODIS. However, the last few years' satellite data of Landsat and Sentinel-2 missions became very promising and the results of yield forecasting on them can be quite good. The important drivers of the 10 and 30 m satellite data usage for the yield forecasting are the implementation of new techniques of data preprocessing that give the possibility to use data collection with higher (2–3 days) temporal resolution rather then 5–16 days and the availability of yield data on higher level—rayon or village level or even field level, rather than oblast or country. A good examples of such research is presented in the work "Winter wheat yield assessment from Landsat 8 and Sentinel-2 data: Incorporating surface reflectance, through phenological fitting, into regression yield models" ([6] and Table 1). In this case it can be proposed to use two data aggregation approaches for the high resolution data for yield forecasting regression fitting. The first is a peak approach. In this case authors using the yield as the dependent variable and the maximum VI's value as independent variables. The second approach is accumulative, when the authors uses the area of curve between the plant emergence date and maximum VI's value achievement date as independent variable. In this case also can be used accumulative degree days (AGDD) instead of accumulative VI. Both approaches demonstrated high accuracy of yield forecasting based on the NDVI, DVI and EVI.

In the case of small number of spatial points, it is better to fit a model for each point based on the high number of temporal distributed data. In this case, we can first calculate the trend line for the yield. The VI's indicators could be used not as predictor of yield. It can be used as predictor for the trend's deviation for the specific year. So the difference between trend and actual yield acts as dependent variable, while the VI acts as independent variable.

**Table 1.** Winter wheat yield model performance ($R^2$, RMSE, RRMSE) on calibration data depending on VI and peak/AUC feature for 2016 to 2018. Best models for each year are marked with italics [6].

| Model | $R^2$ | RMSE, t/ha | RRMSE, % | $p$-Value |
|---|---|---|---|---|
| 2016 | | | | |
| Peak-DVI (data) | 0.179 | 0.308 | 7.7 | $5.61 * 10^{-2}$ |
| Peak-DVI (fitting) | 0.332 | 0.278 | 7.0 | $6.29 * 10^{-3}$ |
| *AUC-DVI* | *0.588* | 0.218 | *5.5* | *$5.02 * 10^{-5}$* |
| Peak-EVI2 (data) | 0.056 | 0.330 | 8.3 | $3.03 * 10^{-1}$ |
| Peak-EVI2 (fitting) | 0.282 | 0.288 | *7.2* | $1.32 * 10^{-2}$ |
| AUC-EVI2 | 0.209 | 0.302 | 7.6 | $3.71 * 10^{-2}$ |
| Peak-NDVI (data) | 0.088 | 0.325 | 8.1 | $1.92 * 10^{-1}$ |
| Peak-NDVI (fitting) | 0.485 | 0.244 | 6.1 | $4.55 * 10^{-4}$ |
| AUC-NDVI | 0.057 | 0.330 | 8.3 | $2.98 * 10^{-1}$ |
| 2017 | | | | |
| Peak-DVI (data) | 0.422 | 0.247 | 7.1 | $2.65 * 10^{-3}$ |
| Peak-DVI (fitting) | 0.400 | 0.252 | 7.2 | $3.67 * 10^{-3}$ |
| *AUC-DVI* | *0.589* | *0.208* | *6.0* | $1.26 * 10^{-4}$ |
| Peak-EVI2 (data) | 0.405 | 0.251 | 7.2 | $3.40 * 10^{-3}$ |
| Peak-EVI2 (fitting) | 0.381 | 0.256 | 7.3 | $4.89 * 10^{-3}$ |
| AUC-EVI2 | 0.570 | 0.213 | 6.1 | $1.87 * 10^{-4}$ |
| Peak-NDVI (data) | 0.388 | 0.254 | 7.3 | $4.38 * 10^{-3}$ |
| Peak-NDVI (fitting) | 0.393 | 0.253 | 7.3 | $4.05 * 10^{-3}$ |
| AUC-NDVI | 0.407 | 0.250 | 7.2 | $3.28 * 10^{-3}$ |
| 2018 | | | | |
| Peak-DVI (data) | 0.597 | 0.176 | 4.7 | $4.53 * 10^{-4}$ |
| Peak-DVI (fitting) | 0.571 | 0.182 | 4.8 | $7.08 * 10^{-4}$ |
| *AUC-DVI* | *0.608* | *0.174* | *4.6* | *$3.66 * 10^{-4}$* |
| Peak-EVI2 (data) | 0.565 | 0.183 | 4.9 | $7.81 * 10^{-4}$ |
| Peak-EVI2 (fitting) | 0.507 | 0.195 | 5.2 | $1.97 * 10^{-3}$ |
| AUC-EVI2 | 0.532 | 0.190 | 5.1 | $1.34 * 10^{-3}$ |
| Peak-NDVI (data) | 0.406 | 0.214 | 5.7 | $7.92 * 10^{-3}$ |
| Peak-NDVI (fitting) | 0.349 | 0.224 | 6.0 | $1.60 * 10^{-2}$ |
| AUC-NDVI | 0.202 | 0.248 | 6.6 | $8.07 * 10^{-2}$ |

However, the best way to build the high quality yield forecasting model is combination of different remote sensing data sources of different types and physics with the adding of in-situ measurements. The world experience in the yield forecasting shows a good informativeness of local agroclimatic data. The weather stations on the fields are the best data sources, but also it is possible to use nearest weather stations and

geospatial data interpolation techniques. Such approach was used for the strawberry yield forecasting on the filed level in the California, USA [7]. The Table 2 shows the weather parameters and adjusted $R^2$ score for respective linear models.

In this way the best regressors were the next:

1. Fall average soil temperature.
2. Net Radiation.
3. Solar Radiation.
4. Cumulated chill hours.
5. Volumetric soil moisture.
6. Soil temperature.
7. Ambient temperature.

A great example of field level yield forecasting using local data shown in [8]. The research was made for maize yield in smallholder farmers' fields in Tanzania. The local data were obtained though the field questionnaire survey in the Survey Solutions. Survey Solutions is a computer-assisted personal interviewing software developed by the World Bank. The trained enumerators administered the field questionnaire survey using tablets with the questionnaire coded in the Survey Solutions application. The enumerators recorded the geographic location and surveyed the physical characteristics of the within-season plant (including planting density, stress level due to N, drought, weeds, pests and diseases) condition. Other in-season information (including weather characteristics and maize cultivar, sowing time, irrigation and fertilization levels) were from enumerators' interviews with the farmers or farm workers. The complete survey was synchronized to the cloud storage. The authors processed the within-season information immediately after they received it through the cloud storage and provided the maize yield forecast for each of the sampling fields. They provided yield forecasts ranging from 14 to 77 days prior to harvest. The 25th and 75th percentile of the forecasting lead time was 30 and 55 days before harvest, respectively.

On Figs. 5 and 6 are shown the reported maize growth conditions on the located fields for three regions. The yield forecasting method is based on this reported data and SALUS crop model [9]. Yield forecasting models are different for these three regions, but the $R^2$ score for them is ranged from 0.5 to 0.94 (Fig. 7).

Data collection in the large scale is not easy task, so it is also possible to combine field level and global datasets using different reanalysis and harmonization approaches. The weather data on filed level provide better performance rather than global products, but the number of fields with weather stations is not so big. Thus, it is possible to combine weather stations to conduct reanalysis of global products and obtained more accurate dataset for yield forecasting [10]. However, using field level data it is possible also run such advance biophysical models, such as WOFOST to obtain high quality crop growth data.

Such models as Crop Growth Modeling System (CGMS) can be run on the geospatial grid with predefined ground date for each point on the large level territory. This model requires the availability of soil profile, daily agro-climatic weather data, list of agromanagement events and crop profile for each node of the grid. It is possible

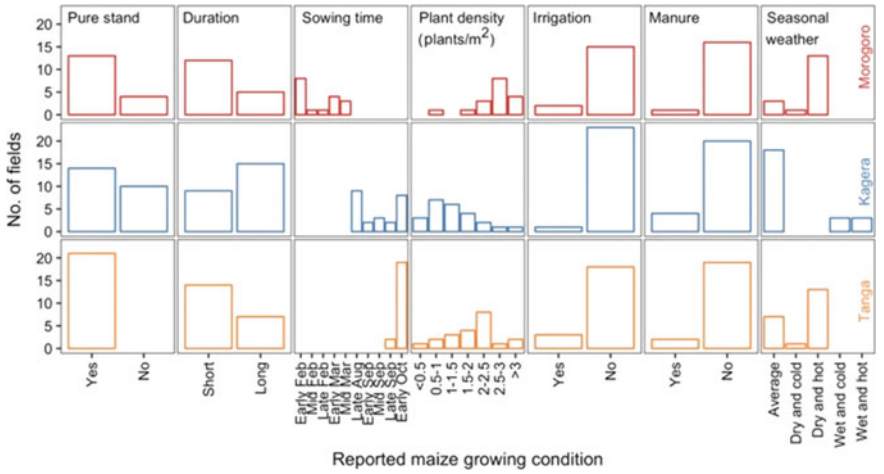



**Fig. 5** Reported maize growing conditions, including pure crop stands versus intercropping, maize duration, sowing time, plant density, irrigation and manure use and growing season weather characteristics across the three districts [8]

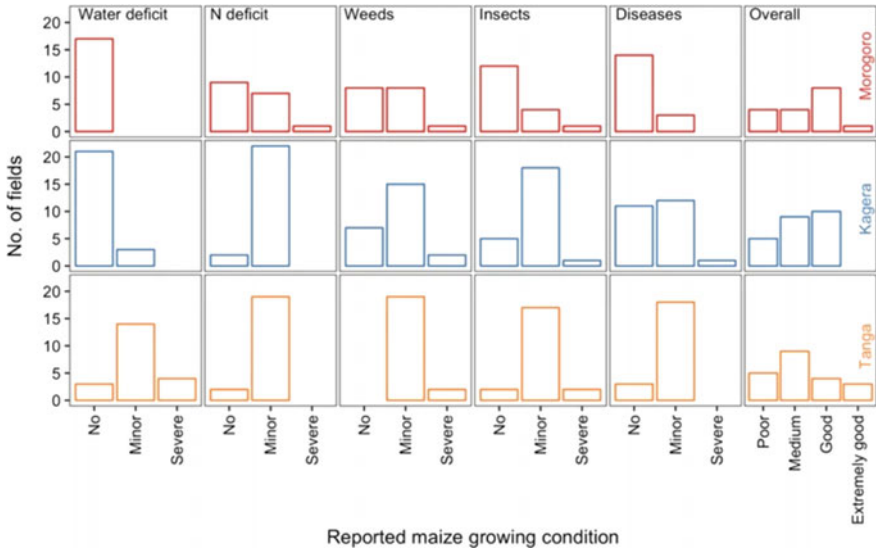



**Fig. 6** Maize status, including water and N deficit, weed, insect and disease presence, and overall plant condition based on photos taken during in-season survey across the three districts [8]

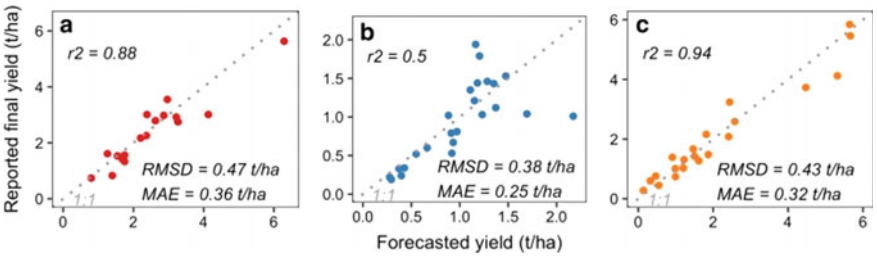



**Fig. 7** Comparisons between the forecasted yield and reported final yield across **a** Morogoro, **b** Kagera and **c** Tanga (note that the ranges for both axes in **a**–**c** differ) [8]

**Table 2.** Weather parameters and adjusted $R^2$ score for respective linear models for yield forecasting fitted on it [7].

| Id | Parameters | Units | Notation | Daily | Moving weekly |
|---|---|---|---|---|---|
| 1 | Average leaf wetness minutes | Minutes | *LWM* | 0.004 | 0.041 |
| 2 | Average leaf wetness count | | *LWC* | 0.120 | 0.274 |
| 3 | Average leaf wetness duration | | *LWD* | 0.092 | 0.148 |
| 4 | Ambient temperature | °C | *ECT1* | 0.460 | 0.505 |
| 5 | Canopy temperature | °C | *ECT2* | 0.030 | 0.116 |
| 6 | Soil temperature | °C | *SMTa* | 0.407 | 0.495 |
| 7 | Volumetric soil moisture | $m^3/m^3$ | *SM* | 0.417 | 0.547 |
| 8 | Daily chill hours | hours | *CHDaily* | 0.189 | 0.292 |
| 9 | Cumulated chill hours | hours | *cumChill* | 0.431 | 0.462 |
| 10 | Reference evapotranspiration | Mm | *ETo* | 0.338 | 0.585 |
| 11 | Solar radiation | $Wm^{-2}$ | *Rs* | 0.421 | 0.667 |
| 12 | Net radiation | $Wm^{-2}$ | *Rn* | 0.439 | 0.656 |
| 13 | Average vapor pressure | kPa | *em* | 0.134 | 0.210 |
| 14 | Average relative humidity | % | *RHm* | 0.000 | 0.002 |
| 15 | Dew point | °C | *dP* | 0.157 | 0.234 |
| 16 | Average wind speed | $ms^{-1}$ | *uBar* | 0.261 | 0.354 |
| 17 | Penmann-Montieth evapotranspiration | mm | *PMETo* | 0.384 | 0.648 |
| 18 | Fall reference evapotranspiration | mm | *ETo.F* | 0.244 | 0.356 |
| 19 | Fall solar radiation | $Wm^{-2}$ | *Rs.F* | 0.129 | 0.196 |
| 20 | Fall net radiation | $Wm^{-2}$ | *Rn.F* | 0.242 | 0.300 |
| 21 | Fall average vapor pressure | kPa | *em.F* | 0.084 | 0.143 |
| 22 | Fall average air temperature | °C | *aTm.F* | 0.270 | 0.449 |
| 23 | Fall average relative humidity | % | *RHm.F* | −0.005 | −0.006 |
| 24 | Fall average wind speed | $ms^{-1}$ | *u.F* | 0.020 | 0.055 |
| 25 | Fall dew point | °C | *dP.F* | 0.071 | 0.116 |
| 26 | Fall average soil temperature | °C | *STm.F* | 0.739 | 0.748 |

to interpolate the model's result from the point measurements to the polygon level if take the polygon's centroids as the model grid points. In this way, these polygons should cover the territory with the same agro-climatic conditions, soil type and agrarian practices. In most cases agromanagement event list will include the same events (planting and harvesting), due to the impossibility to obtain fertilization information on the large level territory, so this list can be one for whole territory. The modeling polygons can be created by intersection of soil map and weather data grid. Also, the important simplification can be done on the use of crop profile. It is possible to use one crop profile for specific crop that was calibrated for the country of interest of climate of interest. If one profile used, the TSUM1 and TSUM2 (sum of active temperature for emerging and sum of active temperature for the flowering) should be calibrated for each point. As the calibration can be used the crop profiles for the regions that cover the territory of interest and weather data. After the running of CGMS, the model the outputs: LAI, total dry weight of roots (TWRT), total dry weight of leaves (TWLV), total dry weight of stems (TWST), total dry weight of storage organs (TWSO), total above ground production (TAGP), harvest index weight of storage organs (HINDEX) can be used for the further regression function fitting for the yield forecasting in the same way as remote sensing data.

## 4 Yield Forecasting Experiment

To conduct yield forecasting, two datasets were prepared. Both are prepared on the base of MODIS MOD13Q1 NDVI product collection, made using Google Earth Engine. The first dataset is the maximal NDVI between 1 of March and 1 of June. It used for the winter wheat yield forecasting. The second dataset is the maximal NDVI between 01 of March and 01 of September. It used for the summer crops yield forecasting.

As the yield forecasting strategy we used the model fitting for each region based on the historical data and yield trend (from 2016 to 2021). In this case we are using one year from 2016 to 2020 as training year, 2021 as validation data. To estimate more accurate model with more appropriate accuracy metric, we iteratively doing the model fitting and 2021 yield forecasting by changing validation year and averaging the final scores outputs. The models fitting was conducted for the winter wheat and maize for Kyiv region. The first step is trend model fitting based on the yield data. It was made iteratively by using yield data subsets with dropping of validation year. The Figs. 8 and 9 shows the averaged trend models for winter wheat and maize. Blue points—yield by statistics, dark blue line—yield—trend, r2 – r squared metrics, MSE1—mean squared error between trend and merged validation + training data, MSE2—mean square error between trend and validation data, MAE1—mean absolute error between trend and merged validation + training data, MAE2—mean absolute error between trend and validation data.

The next step is calculating differences between the actual yield and corresponding trend point for corresponding year. Now these differences can be used for the linear

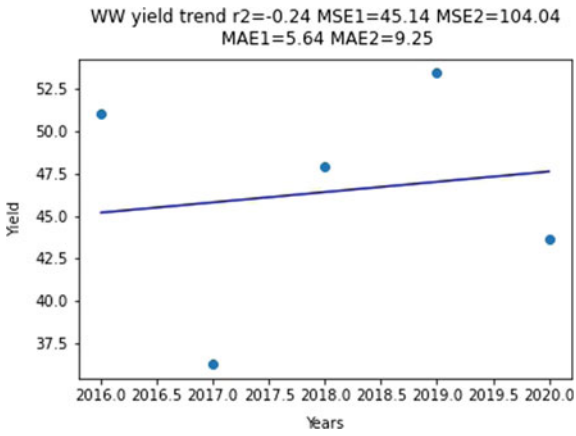



**Fig. 8** Winter wheat yield trend model for Kyiv region

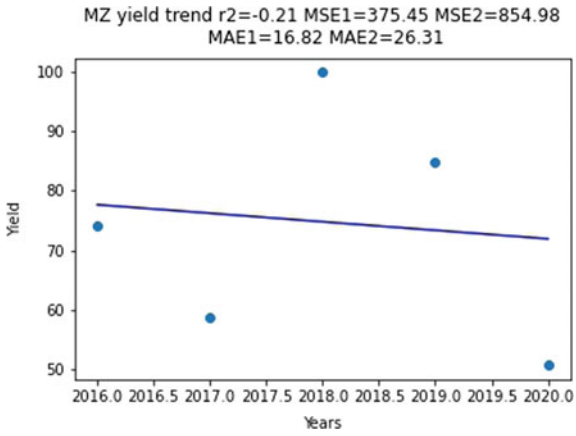



**Fig. 9** Maize yield trend model for Kyiv region

regression function fitting for the yield forecasting. The Figs. 10 and 11 shows the obtained yield forecasting models for the winter wheat and maize. There green points—yield by statistics, red points—predicted yield, dark blue line—yield trend, r2 – r – squared metrics, MSE1—mean squared error between predicted yield and merged validation + training data, MSE2—mean square error between predicted yield and validation data, MAE1—mean absolute error between predicted yield and merged validation + training data, MAE2—mean absolute error between predicted yield and validation data.

The yield forecast for the winter wheat 2021 is 46.81 t/ha, and for maze—75.43 t/ha. The winter wheat has the highest $R^2$ score, which is equal to the 0.67 and low MAE equal 2.6. The maize has much lower $R^2$ score equal to 0.2 and MAE 11.5. The main problem for these models is the significant drop in the 2020 related to the weather conditions and drought. It is especially influenced the maize result, where the difference between actual and predicted yield is almost 29 t/ha.

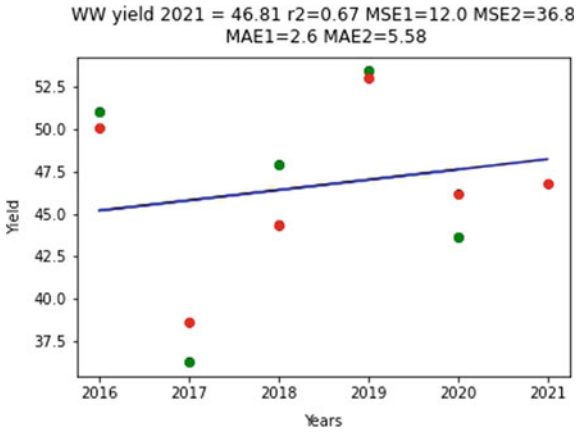



**Fig. 10** Yield forecast for the winter wheat

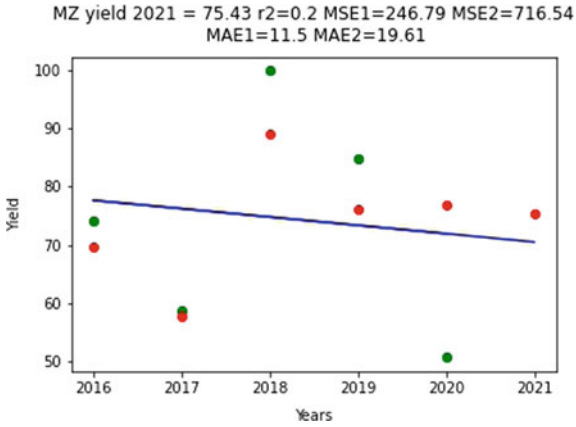



**Fig. 11** Yield forecast for the maize

## 5 Cloud-Based Information Technologies

To conduct our yield prediction experiment, we need to get high-resolution satellite data. The direct download and processing of this data requires high computing resources. It can take a long time to get satellite images, which will significantly slow down the pace. In addition, the problem of big data is closely linked to satellite data, where traditional tools and approaches cannot be applied and highly efficient processing methods are required. And there are difficulties in allocating memory when storing large amounts of data. In particular, Sentinel-2 provides approximately 4 TB of images for the territory of Ukraine, with a cloud cover of less than 20% in one vegetation season. The storage and transmission of large amounts of satellite data could not be effectively addressed without the use of secondary storage. To overcome these problems, it was decided to use cloud technology.

The unique properties of modern cloud infrastructure make it possible to obtain data so quickly that sometimes it takes almost no time to download them. The cloud environment provides the opportunity to use advanced machine methods, as they require a lot of computing resources for effective implementation.

We used cloud infrastructure such as Amazon Web Services and Google Earth Engine to process satellite data. One of the main advantages of this approach is that all satellite data and products are pre-loaded into the cloud and can be used automatically. So we don't need to download the data ourselves.

Technology of Amazon Web Services allows users to have a full-fledged virtual cluster of computers that is always available online. AWS virtual machines have most of the attributes of a real computer, including hardware, optional operating system, network, pre-installed applications such as web server, database, CRM, etc. Each AWS system also virtualizes the console I/O (keyboard, display and mouse) that allows AWS users to connect to their AWS system using a browser. The browser acts as a window into a virtual machine, allowing the user to log in, configure and use their virtual systems just like a real, physical computer. This allows them to configure the system to provide Internet-oriented services and services on the base of standardized predefined AWS services like SageMaker or Kinesis Data Stream. In general, you can use three different infrastructure access modes: via Management Console, Python script or Graphical User Interface (GUI).

In our research, we often take satellite data from the Amazon S3 repository and pre-processed it using Open Data Cube (ODC) software deployed on Amazon EC2. We have some experience to use ODC, which was obtained within different grants. In particular, in this study, we download data from the Sentinel-2 satellite mission from this environment. We use the Open Data Cube for a variety of applications, including analysis of land, water, clouds and time series. Programs for mosaic creation, spectral index calculation, water mapping, land classification, and land change are available after system deployment, and we plan to use them as input for mapping the land cover. An example of a deployed system is shown in Fig. 12.

An important task that Amazon's cloud technology will allow is to pre-process satellite images before they can be used directly. The main problem of satellite changes is cloudiness, the presence of shadows and noise.

The main workflow of satellite data processing is consisting of numerous partial stages. These include orbit correction, boundary noise removal, thermal noise removal, radiometric calibration, orthorectification, filtering, and more. The calculation time depends on the type of instance, but in general, using the usual methods, 1 scene can be processed in 10 min. Optical data require atmospheric correction with masking of clouds and shadows. In total, it takes about 30 min to process one Sentinel-2 scene with Sen2Cor software.

To obtain maps for large areas, you need to register for satellite time series (between images from different sensors and between images in the same time series). The GDAL tool allows to use gdal_merge.py script to perform this task without cloud technology, but it is very time consuming. The problem needs to be solved using the scalable computing infrastructure Amazon EC2. This can be achieved by parallel data processing on multiple computational instances.

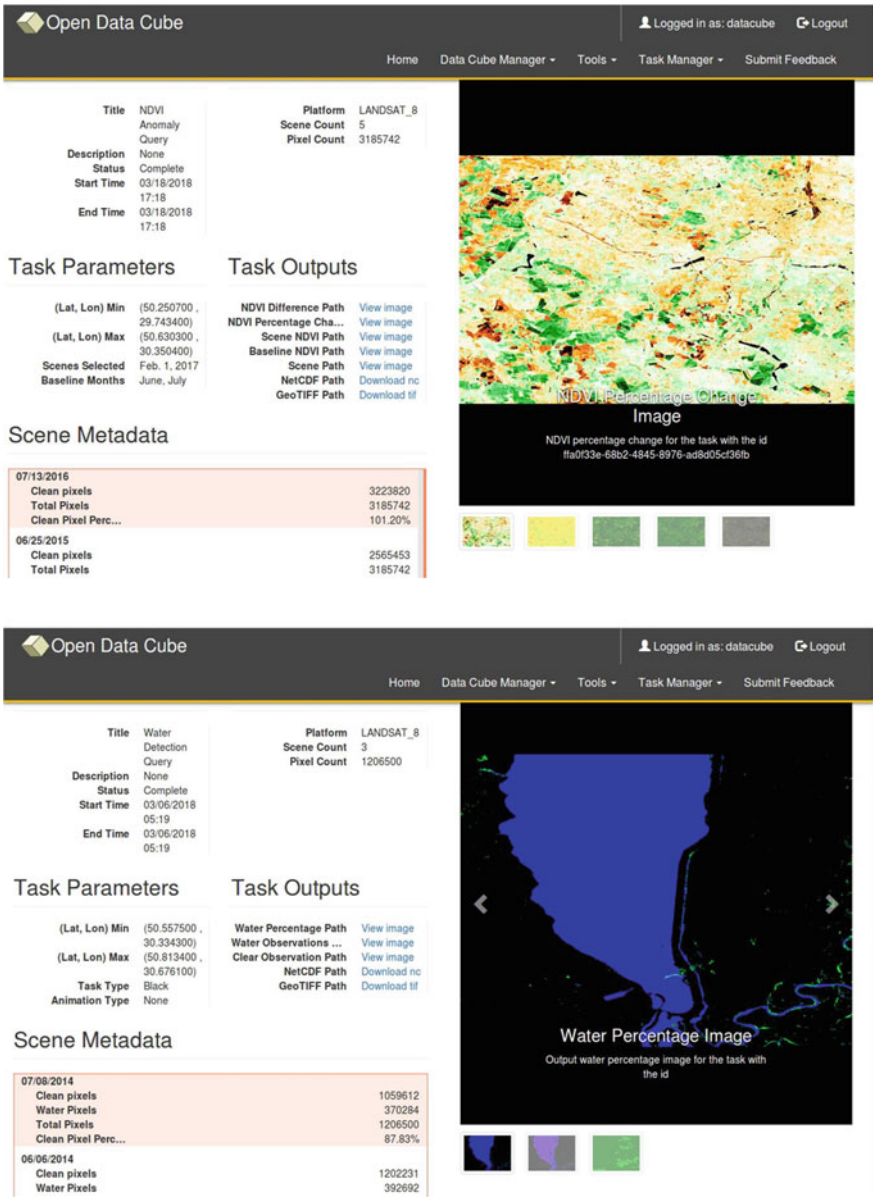



**Fig. 12** Example of deployed Open Data Cube system interface

Google Earth Engine [18] is an another cloud platform for geospatial data analysis on a planetary scale from Google company. It in some sense alternative of AWS services. It leverages Google's vast computing power to study a wide range of issues: forest loss, drought, natural disasters, epidemics, food security, water management, climate change and the environment. Google Earth Engine combines a multi-five-byte catalog of satellite imagery and geospatial data with planetary-scale analysis capabilities, providing built-in features and intra-parallel computing access.

This analysis platform allows to process geospatial information in the cloud without having to use the user's local computer memory. At the same time, GEE is directly connected to several satellite programs that allow to integrate newly captured images into a database. Thus, GEE is great for managing big massive data and allows you to give results quickly, but with predefined set of algorithms. GEE not only uses its own collections, but also allows users to upload data to the environment in raster or vector format. Data processing is performed in the GEE cloud, but there is a feature that allows you to add the created information to the user's Google Drive storage.

Researchers all over the world have long practiced the use of GEE technology for analysis based on remote sensing data. In particular, the platform analyzed the changes in vegetation [19], the dynamics of land use and soil [20], forest transformation from the use of geotechnologies [21], etc.

In this study, we use the opportunities of Google Earth Engine for downloading MODIS satellite images. To do this, we use our own script for downloading data, which indicates the necessary source of data from the platform, the time interval of interest to us, the coordinates of the polygon of the study area or a vector map. The source and download of the data are shown in Figs. 13 and 14.

In addition, cloud technologies make it possible to perform a number of other important tasks of space research. For example, when forecasting yields [2, 11, 12] and solving a number of other tasks, including estimating sown areas [21] and

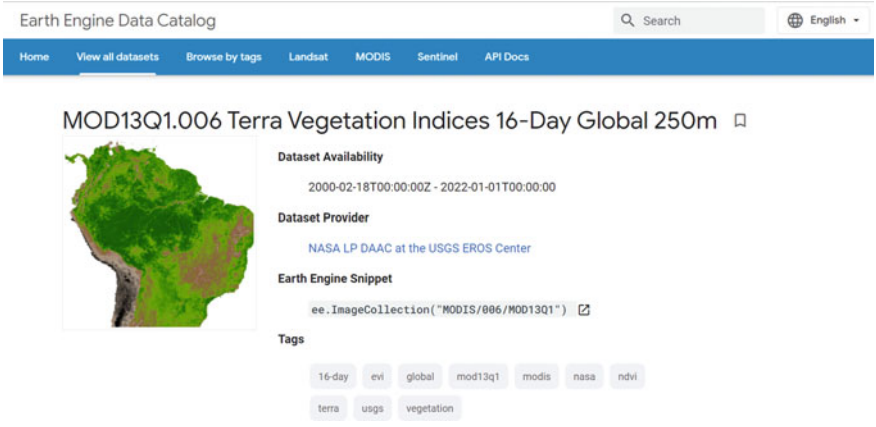



**Fig. 13** Satellite collection MOD13Q1.006 Terry vegetation indices 16-day global 250 m on the platform Google Earth Engine

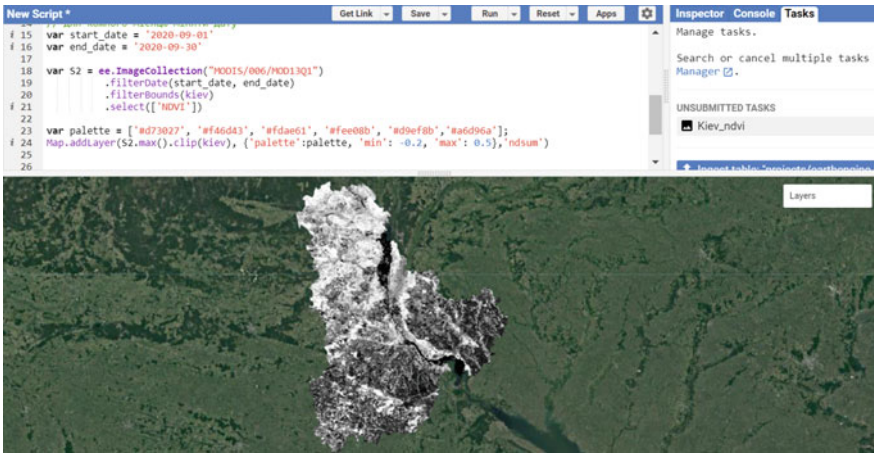



**Fig. 14** Downloading NDVI data for the Kyiv region from the MODIS collection on the platform Google Earth Engine

drought risks [15], mapping crops [14] based on high-resolution satellite data is very important. Cloud platforms make it possible to deploy some crop classification system as depicted in [13] and in Fig. 15.

Here, the great advantage of using cloud innovation is the speed of data processing. Citing an article by the developers of such a system [13], "In the case of a landfill such as Ukraine, the full coverage (9 S1 paths) during the one-year growing season consists of more than 800 scenes (S1A only). Downloading this amount of data from ESA SciHub at relatively low speeds takes a long time (up to 2 full days with an average speed of 5 Mbps with storage in the local infrastructure), while the corresponding download of data for pre-processing on Amazon from Alaska spacecraft will take 10–15 h (with an average speed of 15–20 mb/s)". Thus, the cloud technologies introduced in the system have reduced the access time to remote sensing data by 5–10 times.

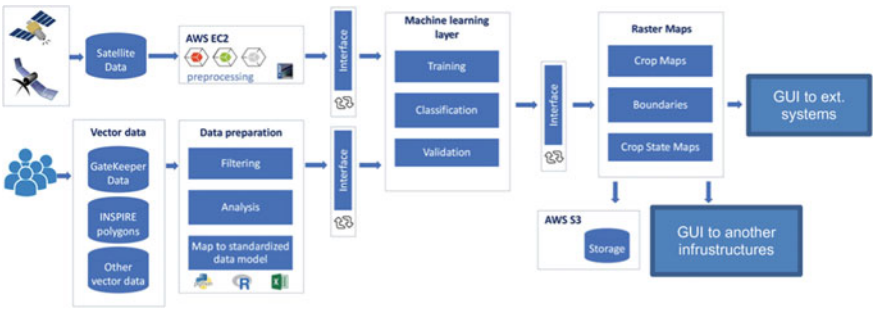



**Fig. 15** Typical architecture component diagram for crop classification system

Thus, the use of cloud technologies allows frequent updates of crop type maps during the growing season with greater accuracy based on time series of satellite images [13]. An example of a classification map, which was built using GEE cloud, is shown in Fig. 16. Accuracy of crop classification during the growing season depending on the number of obtained scenes is shown Fig. 17.

In addition, weather conditions are an important factor influencing land productivity when forecasting yields. Weather simulation is performed through a copy Amazon EC2 using the WRF model. WRF is a numerical weather forecasting system designed for atmospheric research and operational forecasting. Due to the complex modeling process, this system requires a lot of computing resources and a cluster to run the WRF model. We have previously launched the WRF model through the GRID infrastructure [22]. The forecasts were based on a 200 × 200 grid with a cell size of 10 × 10 km. And Amazon EC2 public cloud infrastructure foreign researchers have had successful experience deploying systems for research and weather forecasting [23, 24]. In general, proposed approach on the base of cloud infrastructures and data analysis tools including satellite-based information can be useful for wide range of applied problem, for example in the area of automatic control and dynamic system state estimation [25].

Thus, cloud technologies allow us to quickly download and process numerous collections of satellite images for application. With the help of technologies deployed in the clouds, you can get data on the classification of crops for national use, daily weather data, and so on. Thus, the use of cloud infrastructure solves the problem of processing, downloading and storing remote sensing data of high temporal and

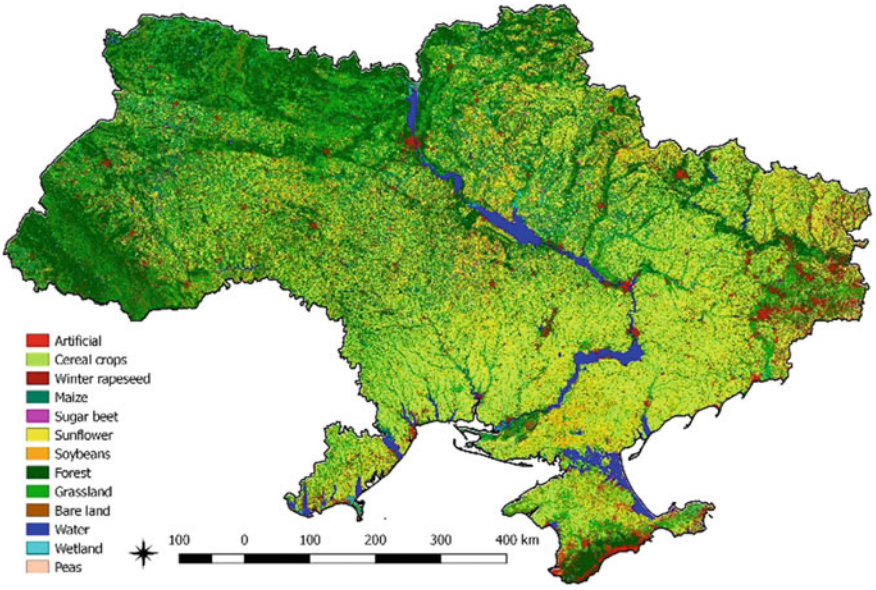



**Fig. 16** Map of crop classification for national use—Ukraine for 2017

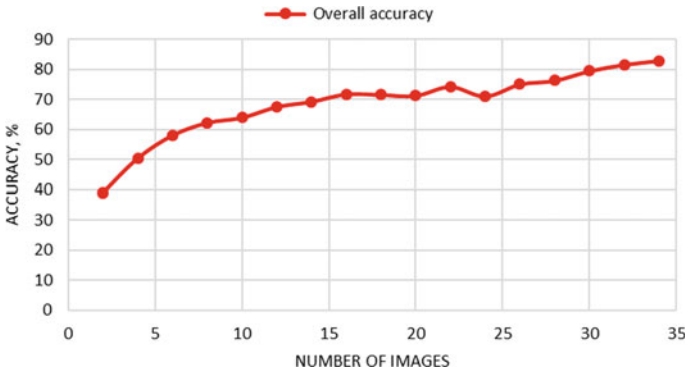



**Fig. 17** Accuracy of crop classification during the growing season depending on the number of obtained scenes (viewing time—6 days)

spatial resolutions, which are so necessary for this study, and in general guarantees the possibility of this work.

## 6 Conclusions

This study showed that Sentinel 2 agro-climatic data, assessment of land degradation on the basis WOFOST model (CGMS with a grid) using AWS and GEE cloud computing platforms is an effective approach in crop yield forecasting. By focusing the 10 m resolution satellite images on wheat fields of MODIS MOD13Q1 NDVI product collection, made using Google Earth Engine the crop forecasting strategy was made by fitting for all Ukrainian regions from 2016 to 2021. Particularly the models fitting was conducted for the winter wheat and maize for the Kyiv region. On the proposed model the crop (wheat) yield forecasting for the winter wheat 2021 is 46.81 t/ha and for maize—75.43 t/ha. All crop yield forecasting results were obtained using cloud infrastructure on the basis of machine learning algorithm and AWS cloud platform.

This study provides objective and promising results for crop yield forecasting using high-resolution Sentinel-2 data, which inherently are big data.

But further study could be focused on:

- Comparing the proposed model predictive power with using different sentinel imagery more accurate data.
- Develop effective methods that allow accounting in-situ in different regions online.

## References


1. Franch B, Vermote EF, Skakun S et al (2019) Remote sensing based yield monitoring: application to winter wheat in United States and Ukraine. Int J Appl Earth Obs Geoinf 76:112–127
2. Kolotii A, Kussul N, Shelestov A et al (2015) Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine. Int Arch Photogram Rem Sens Spat Inf Sci
3. Ma G, Huang J, Wu W et al (2013) Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. Math Comput Model 58(3–4):634–643
4. Mathieu JA, Aires F (2018) Assessment of the agro-climatic indices to improve crop yield forecasting. Agric For Meteorol 253:15–30
5. Skakun S, Franch B, Vermote E et al (2019) The use of landsat 8 and sentinel-2 data and meterological observations for winter wheat yield assessment. In: IGARSS 2019–2019 IEEE international geoscience and remote sensing symposium, pp 6291–6294
6. Skakun S, Vermote E, Franch B et al (2019) Winter wheat yield assessment from landsat 8 and sentinel-2 data: incorporating surface reflectance, through phenological fitting, into regression yield models. Remote Sens 11(15):1768
7. Maskey ML, Pathak TB, Dara SK (2019) Weather based strawberry yield forecasts at field scale using statistical and machine learning models. Atmosphere 10(7):378
8. Liu L, Basso B (2020) Linking field survey with crop modeling to forecast maize yield in smallholder farmers' fields in Tanzania. Food Secur 12(3):537–548
9. Dzotsi KA, Basso B, Jones JW (2013) Development, uncertainty and sensitivity analysis of the simple SALUS crop model in DSSAT. Ecol Model 260:62–76
10. Santamaria-Artigas AE, Franch B, Guillevic P et al (2019) Evaluation of near-surface air temperature from reanalysis over the united states and Ukraine: application to winter wheat yield forecasting. IEEE J Sel Top Appl Earth Obs Remote Sens 12(7):2260–2269
11. Kogan F, Kussul N, Adamenko T et al (2013) Winter wheat yield forecasting in Ukraine based on earth observation, meteorological data and biophysical models. Int J Appl Earth Obs Geoinf 23:192–203
12. Shastry KA, Sanjay HA (2019) Cloud-based agricultural framework for soil classification and crop yield prediction as a service. In: Emerging research in computing, information, communication and applications. Springer, Singapore, pp 685–696
13. Shelestov A, Lavreniuk M, Vasiliev V et al (2019) Cloud approach to automated crop classification using sentinel-1 imagery. IEEE Trans Big Data 6(3):572–582
14. Shelestov A, Lavreniuk M, Kussul N et al (2017) Exploring Google Earth Engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. Front Earth Sci 5:17
15. Skakun S, Kussul N, Shelestov A et al (2016) The use of satellite data for agriculture drought risk quantification in Ukraine. Geomat Nat Haz Risk 7(3):901–917
16. Segarra J, Buchaillot ML, Araus JL et al (2020) Remote sensing for precision agriculture: sentinel-2 improved features and applications. Agronomy 10(5):641
17. Huete A, Justice C, Van Leeuwen W (1999) MODIS vegetation index (MOD13). Algorithm Theoret Basis Doc 3(213):95–309
18. Perilla GA, Mas JF (2020) Google Earth Engine (GEE): a powerful tool that counters the potential of massive data and the efficiency of processing in the Nube. Geogr Invest 101
19. Andrade RLM (2021) Analysis of changes in vegetation from the processing of satellite images on the Google Earth Engine (GEE) platform. Recital-Revista de Educação, Ciência e Tecnologia de Almenara/MG 3(3):48–64
20. Kanya BB, Rosa KKD, Costella RZ (2015) Analysis of the transformation of the Amazon forest from the use of geotechnology: The Google Earth Engine - in elementary school geography lessons. Gaucho Geogr Bull Porto Alegre 42(2):554–568
21. Gallego FJ, Kussul N, Skakun S et al (2014) Efficiency assessment of using satellite data for crop area estimation in Ukraine. Int J Appl Earth Obs Geoinf 29:22–30

22. Kravchenko AN, Kusul NN, Lupian EA et al (2008) Water resource quality monitoring using heterogeneous data and high-performance computations. Cybern Syst Anal 44(4):616–624
23. Vazquez-Poletti JL, Santos-Munoz D, Llorente IM et al (2015) Cloud for clouds: research and weather forecasting in public cloud infrastructure. In: International conference on cloud computing and services science. Springer, Cham, pp 3–11
24. Krishnappa DK, Irwin D, Lyons E, Zink M (2013) CloudCast: cloud computing for short-term weather forecasts. Comput Sci Eng 15(4):30–37
25. Bakan GM, Kussul NN (1996) Fuzzy ellipsoidal filtering algorithm of static object state. Problemy Upravleniya I Informatiki (Avtomatika) 5:77–92

# Compulsoriness and Energy Efficiency in the Decentralized World of Smart Things




**Andriy Luntovskyy, Tim Zobjack, and Mykhailo Klymash**

**Abstract** Energy-efficient, self-sufficient and intelligent networked nodes and IoT devices can be found almost everywhere in everyday life today. These wide range of internet-enabled household and entertainment devices that can be centrally controlled and managed, through medical devices that monitor the personal condition, to the monitoring of entire machine parks and production streets in Industry 4.0 as well as supply chain management issues. Blockchained IoT contributes to the compulsoriness, commitment in the decentralized world of smart things. Secured IoT devices can only be achieved with the combination of known crypto-technologies. Thought a little further, the step-by-step provision of different blockchain-based platforms and solutions can help to reach the desired protection goals.

**Keywords** IoT · Blockchain · PKI · CIDN · Radio and contactless sensing · Energy efficiency · Energy harvesting · Big data analytics · IoT platforms

## 1 The Aims of the Work

1. The given work is aimed to investigation of the energy efficiency issues for IoT devices, based on combined mobile, wireless, sensor and contactless solutions.
2. Furthermore, the security issues are discussed.

A. Luntovskyy (✉)
BA Dresden University of Cooperative Education – Saxon State Study Academy Dresden, Hans-Grundig Str. 25, 01307 Dresden, Germany
e-mail: andriy.luntovskyy@ba-dresden.de

T. Zobjack (✉)
Integration Experts GmbH, Wittenberger Str. 114A, 01277 Dresden, Germany
e-mail: tim.zobjack@integration-experts.de

M. Klymash (✉)
ITRE, Lviv Polytechnic National University, Profesorska Str. 2, Lviv 79013, Ukraine
e-mail: mykhailo.m.klymash@lpnu.ua

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_8

3. The combined energy-efficient solutions are widely interoperable with the 4G/5G base stations and micro-cells, the backboned Wi-Fi access points as well as with inexpensive and energy-efficient RFID/NFC tags.
4. The given scenarios provide multilateral security for IoT devices under use of not only of PKI, hash computing and digital signatures, but also under use of Blockchain technology.
5. Resource-intensive blockchain-based applications are by far compensated via a row of unquestionable advantages, provided by so-called Blockchained IoT.

The remain of the paper is built as follows:

1. Motivation and IoT devices: WSN and further sensing radio/contactless technologies RFID and NFC.
2. Energy efficiency: a holistic approach with "Low Duty Cycle".
3. Security aspects: conventional methods and advanced BC compulsoriness.
4. Case studies.
5. Recent IoT solutions and platforms.
6. Conclusions and outlook.

## *1.1 Motivation and IoT Devices: WSN and Further Sensing Radio Technologies*

The IoT devices use almost the M2M communication style. The following further parameter and distinguishing features are required [1–6, 8–10]:

1. Tiny till measured DR.
2. Good network covering.
3. Interoperability with 4G/5G.
4. Advanced security based on Blockchain and CIDN [7, 11, 12].

Energy efficiency and minimal $CO_2$ footprint are considered as the most important goals for IoT solution development. The following technologies or their combinations are usually deployed (Table 1).

## *1.2 Challenges for IoT: What Does It Mean?*

The paper examines multiple case studies and gives the answers to the key questions about secured and blockchained as well as energy efficient IoT solutions.

IoT features to consider:

1. Small energy consumption and energy efficiency.
2. Wide interoperability to 5G an Beyond (6G), WSN, RFID, NFC, Robotics, Wearable.

**Table 1** IoT sensing technologies (radio/contactless)

| Type | Title | Frequency, MHz | Purpose |
|---|---|---|---|
| Radio sensing | | | |
| WAN | SigFox | 868 MHz (EU), 902 MHz (USA) | Low-power wide-area network |
| | LoRa | 2,4 GHz, 868/915 MHz, 433 MHz, 169 MHz | Low-power wide-area network |
| | NarrowBand-IoT | 700–2100 MHz | 4G mobile network |
| | LTE-CatM | 700–2100 MHz | 4G mobile network |
| LAN | WiFi5 IEEE 802.11ac | ISM 2400/5000 MHz | Wireless LAN |
| | WiFi6 IEEE 802.11ax | ISM 2400/5000 MHz | Wireless LAN |
| PAN | ZigBee IEEE 802.15.4 | 900/2400 MHz | Low-power personal-area network |
| | 6LoWPAN | ISM 2400/5000 MHz | Low-power personal-area network |
| | Bluetooth IEEE 802.15.1 | ISM 2400/5000 MHz | Low-power personal-area network |
| | EnOcean | 868 MHz (EU), 928 MHz (Japan), 902 MHz (USA) | Low-power personal-area network (Energy Harvesting) |
| Contactless sensing | | | |
| Tags, labels, chipcards | RFID | LW 125–134 kHz, SW 13,56 MHz; UHF 865–869 MHz (EU), UHF 950 MHz (US/Asia), SHF 2,45 GHz/5,8 GHz | Multiple standards |
| | NFC | 13,56 MHz | Standardized RFID, ISO 14443/ISO 15693 |

3. New efficient communications models with decentralization, M2M communication style, fog-based models, P2P instead of convenient C-S, cloud-centric and fog-decentralized.

However, the slogan is as follows: "Energy Efficient and Blockchained IoT" (refer Fig. 1).

The expenditures for IoT & Blockchain worldwide reached together more than 17% of worldwide IT expenditures by year 2018 and 26% by year 2022 [4–8]. Among them 0,03% of worldwide IT expenditures by year 2018 and 0,32% by year 2022 for Blockchain deployment. The IT expenditures yearly reach last 5 years more than 3500 in billions USD.

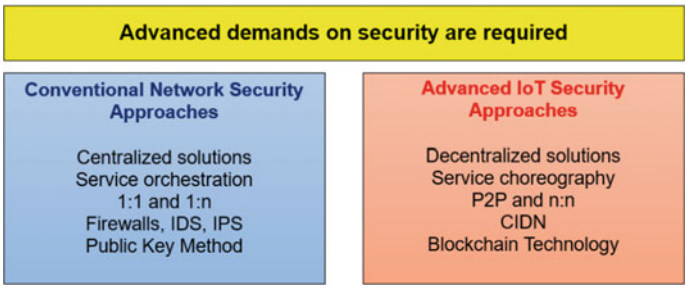



**Fig. 1** Challenges for IoT

Energy efficiency via so-called "Low Duty Cycle" principle is provided under support the holistic multilayered approach [4, 6, 11]. Wireless sensor piconets are, in general, a subject to the condition of unified development with optimization of energy consumption in all layers of the OSI reference model up-to-down and across. The uniform approach to energy efficiency through so-called "Low-Duty-Cycle" is followed [1, 2, 6]. The minimal energy consumption is foreseen for only the modes "Idle" and "Sleep". The ratio a for duty percentage, refer (1), is represented:

$$\begin{aligned} a &= Tduty/Toverall \\ a &= Tduty/(Tduty + Tsleep) \end{aligned} \tag{1}$$

i.e. high energy consumption is typical for "Transmitting", "Receiving" as well as "Normal": "Computing", "Transition" … Mostly sleeping sensors are then energy-efficient. The reference value for the Duty Cycle a is about: 7–10%.

## 2 Advanced Security with BC

The conventional security is provided with PKI. The scheme for the general Public Key Method is depicted in Fig. 2. The assignment of the protection goals [11, 12] to the security mechanisms known for IoT solutions is given below (Table 2).

So-called Advanced Security for IoT devices can be provided via Blockchain (BC) technology, as a continuously expandable list of data records (hashes and transactions) in individual chained blocks [4–7] aimed to secure transactions processing.

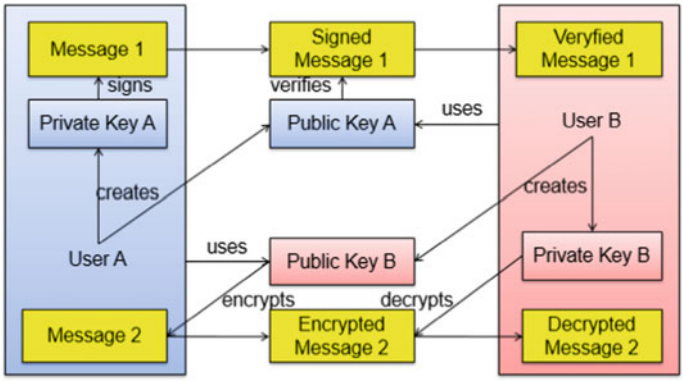



**Fig. 2** General public key method

**Table 2** Protection goals and suitable security mechanisms

| Protection goal | Security mechanism | Example |
|---|---|---|
| Confidence | Encryption | AES |
| Compulsoriness | PKI or Blockchain | RSA, Hash |
| Authorized resource access | Authentication | Login |
| Integrity | Secured data, message authentication code | CRC, Hash |
| Blocking of unauthorized accesses to the networks | Firewalls and collaborative intrusion detection | Firewalls, IDS, CIDN |

Modern IoT applications and mobile apps are nowadays mostly very complex, and consist of multiple communicating parts. BC in combination with CIDN provide better security for IoT solutions [4, 7, 11]. The potential benefits of so-called "Blockchained Apps":

- Software solutions decentralization.
- Compulsoriness of the workflow steps for critical applications.
- Multilateral communication models like e.g. M2M (machine-to-machine).

A typical scenario for "a blockchained mobile network" was discussed in [8]. The required compulsoriness for 5G and Beyond slicing and handover scenarios can be supported via a private BC (Fig. 3) with the involvement of Smart Contracting [10–12].

A typical scenario for "blockchained IoT" is a Supply Chain for goods, that need constant surveillance. The blockchain can be used to store data from sensors in a secure, transparent and near temper-proof manner. It can be used within a company as well as in a collaboration of multiple companies [4–7]. The workflow step compulsoriness for multilateral communication is here required.

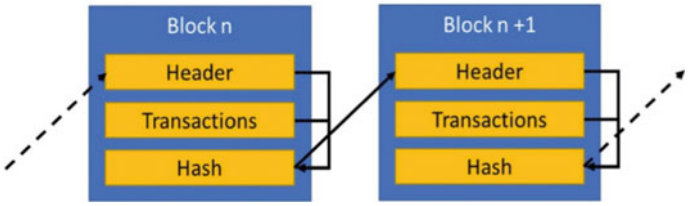



**Fig. 3** Advanced security with BC (simplified)

## *2.1 Blockchain*

The blockchain is a decentralized network, that consists of interconnected nodes in which data and information is redundantly stored. Behind every node is usually a unique participant of the network. Transactions between participants are always directly carried out, like in a Peer-2-Peer network, an intermediary is not necessary. As the name suggests, the blockchain consists of blocks that are connected in numerical order. Each of the blocks has a collection of transactions and metainformations. As soon as a block has reached its maximum number of transactions, all its content is used to generate a hash-code. This hash is then used in the metainformation of the next block and so on. This leads to a connection-chain for all blocks within a network (refer Fig. 3).

Every node in the network has a copy of the complete blockchain and the hash of every block can be checked for its validity. When a transaction in a block is altered, the newly generated hash-code will not match the existing one, the same applies to all blocks that follow.

There are different options, how a blockchain can be set up. It can be public, so everyone can participate by creating its own node, or it can be private, where only a few parties and nodes are present.

With a blockchain, different kinds of security aspects can be achieved:

- Authentication: Signature and encryption can be used, as mentioned before.
- Access Control: In a private blockchain, granular access policies can be defined for different participants.
- Confidentiality: Data is only accessible from within the network.
- Integrity: Transactions that are saved in a block cannot be deleted and are nearly temper-proof.
- Transparency: Every participant has access to the complete blockchain.

## *2.2 Supply Chain Management*

BC supports compulsoriness of the workflow for so-called “supply chains”, i.e. Supply Chain Management in industries, agriculture, logistics, shipping, trading

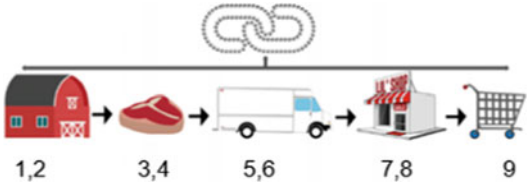



**Fig. 4** Supply chain management based on BC

and finances [6, 7, 11]. Exemplarily Supply Chain Management for meat production, delivery, purchase and consumption, based on BC, is represented below (refer Fig. 4). The "blockchained" steps for a meat quality workflow are as follows:

1. Cattle feeding data.
2. Position data of the cattle, which provide information on keeping.
3. Obtaining verified data on the meat quality.
4. Automatic selection of a logistics solution.
5. Proof of compliance with the cold chain.
6. Optimization possibility in the delivery process.
7. Has transparency about the time of delivery.
8. Automatic adjustment of the order quantity.
9. The customer can use a QR code to access tamper-proof data about the goods.

## *2.3 Smart Contracting*

Smart contracting (SC) provides more security via combined AES/RSA/Hash method what is underlying for BC and has the following further advantages:

- Decentralized processing of contracts based on Blockchain.
- Mapping of the contracts as executable source code.
- Compulsoriness and trustworthiness through transparency.
- "Open Execution" instead of just "Open Source".
- Legal security without an intermediary (jurist).

Figure 5 compares the both approach: (a) Conventional Contracting Proof vs. (b) Smart Contracting.

A Smart Contracting workflow is shown in Fig. 6. A private BC enables the compulsoriness by execution of the workflow steps and can be used for support of IoT devices and apps.

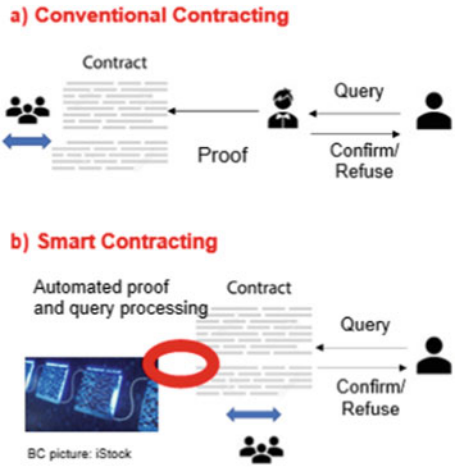



**Fig. 5** Automated proof and query processing by SC

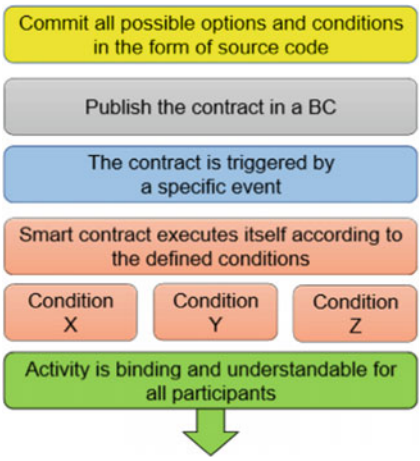



**Fig. 6** A smart contracting workflow

## 3 Case Studies and Best Practices

The case studies and best practices are provided in [4, 7, 11, 12], inter alia:

1. Monitoring and management via the combined RFID/Wi-Fi solutions, BC-based Supply Chain Management.
2. Star without gateway (GW): NB-IoT, LTE-CatM.
3. Star with GW: EnOcean.
4. Mesh with GW: EnOcean.
5. LoRa WAN with GW.
6. Microcell-based energy-efficient hierarchical WSN/Wi-Fi system.
7. GW-based WSN/Wi-Fi with layered infrastructure.

8. Annual Costs for WSN or Wi-Fi etc.
9. One of them is mostly interesting on the scope of energy efficiency.

## *3.1 Scenario 1: RFID and Wi-Fi Based Monitoring and Management of Farm Animals/Meat Cattle*

Mobile native and Web-Apps supports the monitoring of the stall layout, cattle tracking, feeding, cow milking and growth, as well as the alarms due to animal health status (Fig. 7). The monitoring and management of the animal farm/meat cattle is based on RFID and Wi-Fi [6]. A meat quality workflow is based on Blockchained supply chains (s. previous section).

All data along the supply-chain gets aggregated within the blockchain, whether it comes from sensors, manual or automated entries. With pre-defined policies for all participants the access to specific data can be granted and restricted, so everyone can get exactly the data, that is needed. Complete trust between the participants is no longer a necessity, the information in the blockchain can be checked anytime from anyone within the network.

The temper-proof nature of the blockchain also helps with compliance within a company and allows the involvement of government agencies to reliably monitor the supply-chain where it required.

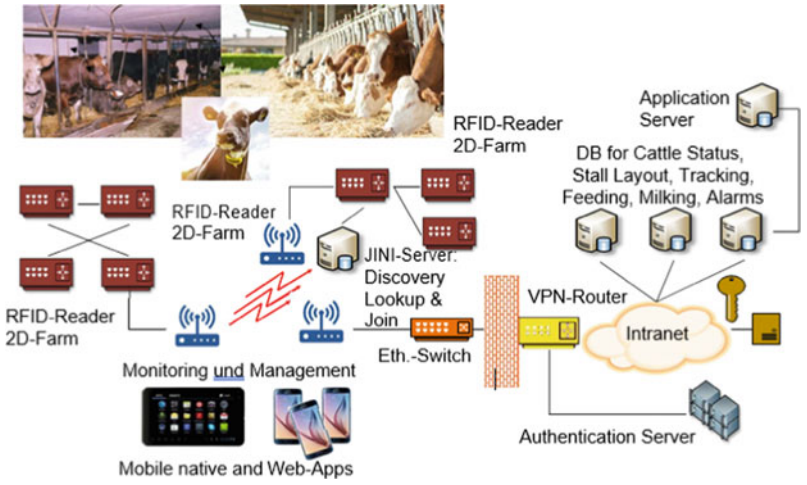



**Fig. 7** RFID and Wi-Fi based monitoring and management of farm animals/meat cattle

### 3.2 Scenario 2: Energy Efficient Sensor Constellation

There are different topologies possible (refer Table 1):

- EnOcean star with GW and Energy Harvesting with the frequencies by F = 868 MHz and 2400 MHz.
- EnOcean mesh with GW and Energy Harvesting with the frequencies by F = 700–2100 MHz.
- Star without GW: NB-IoT, LTE-CatM.

There is a next energy-efficient scenario (Fig. 8). The sensor constellation is considering for the best positions for solar energy harvesting. The mesh of EnOcean sensors communicate with the One Hope GW and then with the cloud with dedicated application server for data acquisition, processing and DB retrieving.

The favoring factors for EnOcean and NB-IoT constellation both are grouped to Table 3.

Table 4 provides a further comparison between NB-IoT, LTE-CatM and EnOcean sensors.

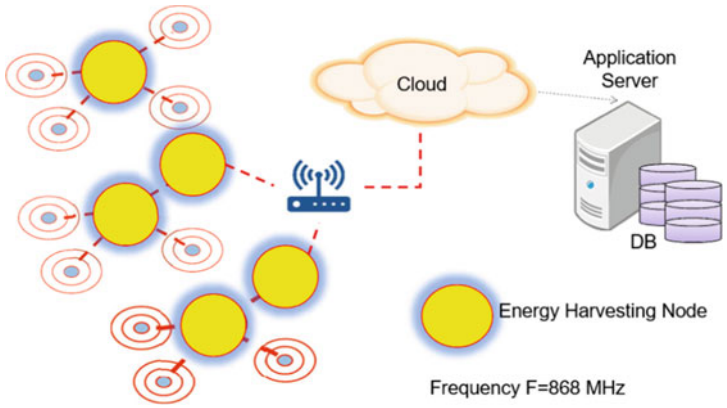



**Fig. 8** Mesh with GW based on EnOcean

**Table 3** The favouring factors for EnOcean and NB-IoT constellation [4]

| The favoring factors for EnOcean | Favoring factors for NB-IoT |
|---|---|
| • Star and mesh topology with Energy Harvesting<br>• Large number of sensors within a small area<br>• Time-critical applications, such as a workplace booking system<br>• Medium to long-term planned use of the sensor network<br>• Square office complexes without tubular high building sections | • Star Topology for NB-IoT<br>• Small number of sensors within a large area<br>• Applications that do not require a live view<br>• Requirement for the rapid establishment of a sensor network<br>• Low transmission frequency of the sensors, Low Duty Cycle |

**Table 4** Energy supply and energy harvesting [4]

| WSN | NB-IoT | LTE-CatM | EnOcean |
|---|---|---|---|
| Idle mode | 0.2 mA × 5 Min | 0.01 mA × 5 Min | 0.001 mA × 5 Min |
| Sense-Transmit-Mode | 20 mA × 0.2 Min | 35 mA × 0.27 Min | 5 mA × 0.08 Min |
| Battery durability | 0,6 years | 0,6 years | 3 years |

### *3.3 Scenario 3: Annual Costs Calculation and Accumulated Data*

Let us a ZigBee WSN to consider which consists of 10.000 sensors (refer Table 1). Depending on the event frequency within this WSN, the event driven, periodically, permanent, and the behavior of the sensors (push, pull), a sensor survey and data accumulation are performed. Thus, if each sensor can transfer a short telegram up to 100 Bits only, then:

- The survey for each sensor is conducted 20 times per hour: 2000 Bits/h.
- × 24 h = 48000 Bits/daily = 6000 Bytes/daily for each sensor.
- In general, an average sensor wireless network accumulates experimental data for 3 years × 365 days: × 365 × 3 = 6.57 MB for each of the sensors.
- The overall-data for the mentioned network: 6.57 MB × 10.000 sensors ~ 66 GB of raw data.

The WSN with One-hope-GW (as alternative to Wi-Fi6) is used. The system is energy-efficient and provides low data unit costs for the collected data. The reference data for ZigBee (IEEE 802.15.4) are as follows:

- Frequency F = 900/2400 MHz (ZigBee).
- Period T = 3 years for N = 10.000 sensors.
- Telegram FL = 100 Bit.
- Data rate DR = 2000 Bits/h (event frequency depended).
- Costs per kWh = 29ct/kWh (Note: average electricity price in Germany).
- Enabled Transmitting Power PTx = 1 mW (0 dBm).
- Average Electricity Price: AEP = 29 ct/kWh.
- Calculation of electricity costs = 1 mW * 24 h * 365 d = 8.76 Wh = 8.76 * 10 ** −3 kWh.
- Amount [ct] = 8.76 * 10 ** −3 kWh * 29ct/kWh = 0.254ct/year.
- Overall Electricity Costs for 10.000 sensors: C = 25 EUR 40 ct.
- Annual Data Unit Costs: ADUC = 38 ct/GByte.

Then, within the 3 years the electricity costs can only reach 76 EUR 20 Cent. The unit costs per GByte for the acquired data in the given case are tiny and amounts to: 76,2 EUR/66 GByte = 1,15 EUR/GByte only (i.e. 38 ct/GByte yearly).

In comparison to ZigBee is the deployment Wi-Fi 6 for 100–200 times more expensive (because of Enabled Transmitting Power is 20 dBm or, even, 23 dBm).

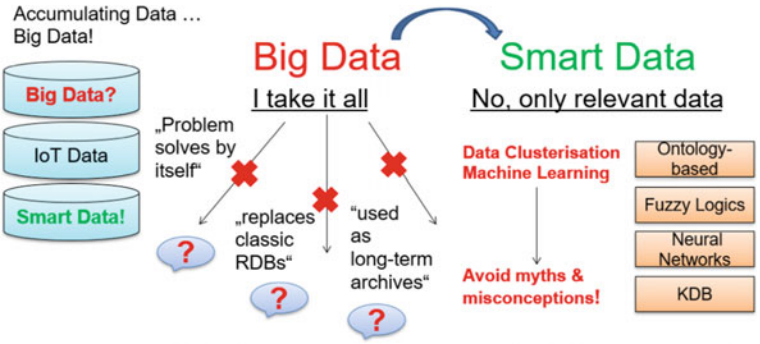



**Fig. 9** “Big Data” to “Smart Data” transition

It means that the energy efficiency is 100 times or, even, 200 times higher for the WSN in comparison to Wi-Fi 6 (IEEE 802.11 ax) [6].

However, the IoT data accumulating can lead to so-called big data [9] problematics practically inevitably (Fig. 9).

Smart data for IoT devices is of great importance (Fig. 10). Aimed to avoiding myths and misconceptions such as [9]:

- The problem solves itself.
- Such data replace only the classic relational DB.
- Such data can be used as long-term archives, the data must be converted into smart data, which is also a challenge here. The wrong developer position “I take it all” must be replaced by “No, only relevant data”. This can be achieved through a preliminary “data clusterization” and, then ML processes. From this point of view, the following methods are usually considered [9]: ontologies, fuzzy logic, neural networks, knowledge DB.

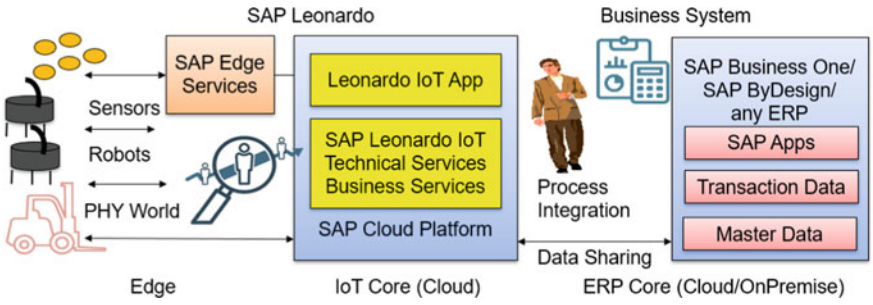



**Fig. 10** SAP based IoT integration

## 4 Recent IoT Solutions and Platforms

Thought a little further, the step-by-step provision of different blockchain-based platforms and solutions can help to reach the desired energy and data efficiency as well as the protection goals (Fig. 10). There is nowadays a large number of platforms from well-known software manufacturers for the management and data processing of IoT devices [5, 7]:

- IBM Watson IoT Platform.
- Microsoft Azure IoT Hub and Win for IoT.
- Google Cloud IoT.
- Amazon AWS IoT.
- SAP Internet of Things, SAP Leonardo IoT, SAP Edge Service (refer Fig. 10).

On the other hand, multiple open-source SW solutions and platforms for IoT device integration can be mentioned like Robot OS, OPC UA, RabbitMQ, MosQuitto, AutomationML tools, which are based the both on the above listed application protocols (refer Table 1). These can be taxonomized by their universality and the supported communication protocols: from the mostly simple tools and frameworks up to the whole integration platforms.

A similar picture can be drawn for blockchain-platforms. More and more hyperscalers bring their own cloud solutions:

- IBM Blockchain.
- Microsoft Azure Blockchain.
- Oracle Blockchain Cloud Service.

They can be used as entry to connect with IoT and other enterprise solutions to create complete business processes.

Besides the enterprise-grade platforms, there are also the existing blockchain technologies like Ethereum, Hyperledger Fabric, Quorum, Stellar, Ripple, Bitcoin, and others, with their corresponding platforms. But they are mainly used as open networks, outside of business scopes [13].

## 5 Conclusions and Outlook

1. The work represents a short overview on WSN and further sensing radio technologies as well as contactless technologies RFID and NFC.
2. These technologies are suitable for the deployment by IoT, IIoT and Smart Office as well as aimed to support of energy-efficient and secured systems.
3. Additionally, they are inter-operable with 5G and Beyond, like NB-IoT, LTE CatM, LoRa WAN, Wi-Fi 5/6.
4. Energy efficiency is considered here as one of the most important issues for IoT.

5. A holistic approach for energy efficiency for IoT is provided on the basis of "Low Duty Cycle" principle for the deployed sensors. Some scenarios illustrated the above-mentioned approach.
6. Furthermore, the security aspects for IoT are examined: conventional PKI and advanced "blockchained" compulsoriness are compared.
7. The given scenarios provide multilateral security for IoT devices under use of not only of hash computing and digital signatures, but also under use of Blockchain.
8. Resource-intensive blockchain-based applications are by far compensated via a row of unquestionable advantages, provided by "blockchained IoT".

**Acknowledgements** Authors' great acknowledgements to the colleagues from BA Dresden (Saxon Academy of Studies) and Lviv National Polytechnic University, especially to Dr. habil. T. Maksymyuk, PhD student B. Shubyn, BA students T. Schwabe, A. Podoprygora, O. Graetsch, M. Stoll for support, inspiration and challenges by fulfilling of this work.

## References

1. Energy Harvesting. https://powunity.com/lte-m-die-neue-form-der-iot-technologie
2. EnOcean. https://www.enocean-alliance.org
3. Zaigham M (ed) (2017) Fog computing: concepts, frameworks and technologies. Springer, London. ISBN 978-3-319-94890-4
4. Luntovskyy A, Zobjack T, Shubyn B, Klymash M (2021) Energy efficiency and security for IoT scenarios via WSN, RFID and NFC. In: IEEE 5th international conference on information and telecommunication technologies and radio electronics (UkrMiCo-2021), Kyiv, Ukraine, 29 November–3 December 2021. NTUU "KPI Igor Sikorsky", Kyiv, pp 1–6
5. Zobjack T, Luntovskyy A (2020) Blockchained IoT: Verbindlichkeit in der dezentralisierten Welt smarter Dinge, 4. Wissen im Markt, German, pp 75–81. ISSN 2512-4366, https://www.ba-sachsen.de
6. Luntovskyy A, Shubyn B (2020) Energy efficiency for IoT, based on IEEE workshop RECI-2020 on reliability engineering and computational intelligence, 27–29 October 2020 in Springer LNCS series reliability engineering and computational intelligence - studies comp. intelligence by Gulijk C, Zaitseva E (eds), July 2021 vol 976, pp 199–215. ISBN 978-3-030-74555-4
7. Luntovskyy A, Zobjack T (2021) Secured and block chained IoT, IDT-2021 Zilina, IEEE Xplore, 22–24 June 2021. In: IEEE issue "information and digital technologies, pp 1–10. ISBN 978-1-6654-3692-2, ISSN 2575-677X
8. Luntovskyy A, Shubyn B, Maksymyuk T, Klymash, M.: 5G slicing and handover scenarios: compulsoriness and machine learning. In: Vorobiyenko P, Ilchenko M, Strelkovska I (eds) Current trends in communication and information technologies. Springer lecture notes in networks and systems series, based on Conf. IPF-2020 "Infocommunications - present and future", Odessa, 16–19 November 2020, vol 212, pp 223–255. ISBN 978-3-030-76342-8
9. Luntovskyy A, Globa L, Shubyn B (2020) From big data to smart data: the most effective approaches for data analytics. In: Advances in information and communication technology and systems, Part of the "Lecture notes in networks and systems series", MCT 2019. LNNS, vol 152. Springer, pp 23–40. ISBN 978-3-030-58359-0. https://doi.org/10.1007/978-3-030-58359-0_2
10. Luntovskyy A, Beshley M, Klymash M (eds) (2022) Future intent-based networking: on the QoS robust and energy efficient heterogeneous software-defined networks. In: LNEE, vol 831. Springer, 530 p. ISBN 978-3-030-92433-1

11. Luntovskyy A, Guetter D (2022) Highly-distributed systems: IoT, robotics, mobile apps, energy efficiency, security. Springer Nature Vieweg, 350 p. ISBN 978-3-030-92828-5
12. Luntovskyy A, Beshley M (2022) Designing HDS under considering of QoS robustness and security for heterogeneous IBN. In: Luntovskyy A, Beshley M, Klymash M (eds) Springer LNEE, vol 831, Springer, pp 19–37. ISBN 978-3-030-92433-1
13. Blockchain platforms reviews and ratings. https://www.gartner.com/reviews/market/blockchain-platforms

# Standard Model of System Architecture of Enterprise IT Infrastructure




**Stanislav Dovgyi and Oleh Kopiika**

**Abstract** This paper proposes a modern IT infrastructure model for an enterprise with own Data Centers and geographically dispersed organizational structure. The model includes a set of architectures, namely security, management, data storage, applications, and network. Architectures define the fundamental principles for building IT services and their relationships. Additionally, each architecture is used as the basis for setting up the requirements for the creation of IT services. The suggested principle for developing a typical system IT infrastructure is as follows: architectures define a set of services; IT services are provided to three groups of customers; IT services and customers are interrelated within five implementation scenarios; five architectures define the integration of IT services. The paper proposes to use the information technologies as IT Services, which aim at maintaining the following elements in a technically sound condition: network devices, computing equipment, data storage devices, automatic software deployment service, network service, perimeter security service, directory service, file and print service, data management service, business application service, IT management service, backup and recovery service, certificate management service, integration service. All customers of the Enterprise are divided into three main categories. Where appropriate, the customers may be further grouped within each category as follows: employees, partners and partner organizations, external customers. Implementation scenarios: Data Centers, department, remote office, Extranet, Internet Data Center. A method of optimizing the process of providing for certain categories of clients with Data Center services is proposed for mathematical modeling of the corporation's IT infrastructure, which brings the solution to the incidence matrix for certain graph. As an example of designing elements of IT infrastructure, the design of optimal network architecture is considered. Examples of implementation of the proposed methodology for building IT infrastructure for service centers of PJSC "Ukrtelecom" are given.

S. Dovgyi (✉) · O. Kopiika (✉)
Institute of Telecommunications and Global Information Space of NASU, Chokolivskiy Blvd. 13, Kyiv 03186, Ukraine
e-mail: s.dovgii@gmail.com

O. Kopiika
e-mail: okopiyka@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_9

**Keywords** IT infrastructure · Data center · IT services · IT infrastructure architecture

## 1 Introduction

The main objective of this study is a building IT infrastructure for complex automated control systems of the manufacturing, operating activity, production facilities, system-wide support, as well as tools that allow you to create, process, store, control, monitor, diagnose and transport information using a unified system of business processes [1–7].

Technology is at the heart of virtually every process at a modern enterprise, from employee work activity management to day-to-day operations, manufacture of goods, and provision of services.

A flexible, reliable and secure IT infrastructure helps the Enterprise achieve its goals and gain a competitive advantage in the market. However, mistakes made during the implementation of IT infrastructure can lead to interoperability, performance, and security issues, including systemic failures and data leakages. Properly implemented infrastructure may be described as a key business profitability factor.

IT infrastructure is a set of interrelated information systems and services, which ensure the functioning and development of Enterprise information interaction tools [8].

According to the ITIL Foundation Course Glossary, IT Infrastructure can also be termed as "All of the hardware, software, networks, facilities, etc., that are required to develop, test, deliver, monitor, control or support IT services. The term IT infrastructure includes all of the Information Technology but not the associated People, Processes and documentation" [9].

Leaders and managers within the IT field are responsible for ensuring that both the physical hardware and software networks and resources are working optimally. IT infrastructure can be looked at as the foundation of an organization's technology systems, thereby playing an integral part of driving its success [10].

With the current speed that technology changes and the competitive nature of businesses, IT leaders have to ensure that their IT Infrastructure is designed such that changes can be made quickly and without impacting the business continuity [11].

While traditionally companies used to typically rely on physical Data Centers or colocation facilities to support their IT infrastructure, cloud hosting has become more popular as it is easier to manage and scale. IT Infrastructure can be managed by the company themselves or it can be outsourced to another company who has consulting expertise to develop robust infrastructures for an organization [12].

Regardless of the chosen IT infrastructure type, it is important to formulate a development strategy for the Enterprise's IT infrastructure based on advanced methods and concepts of leading hardware and software manufacturers.

## 2 Formulation of the Problem

The IT infrastructure of an Enterprise should meet the following requirements (see Fig. 1) [13–18]:

1. High availability of IT services.
2. High security.
3. Scalability of IT infrastructure and individual components.
4. Manageability.
5. Support.
6. Replicated solutions.
7. Standardization.
8. Integration.
9. Potential for upgrading.

To build the IT infrastructure of an Enterprise, we will deal with the following four components:

- *Ss*—IT Services.
- *K*—Customers.
- *Sc*—Scenarios.
- *A*—Architectures.

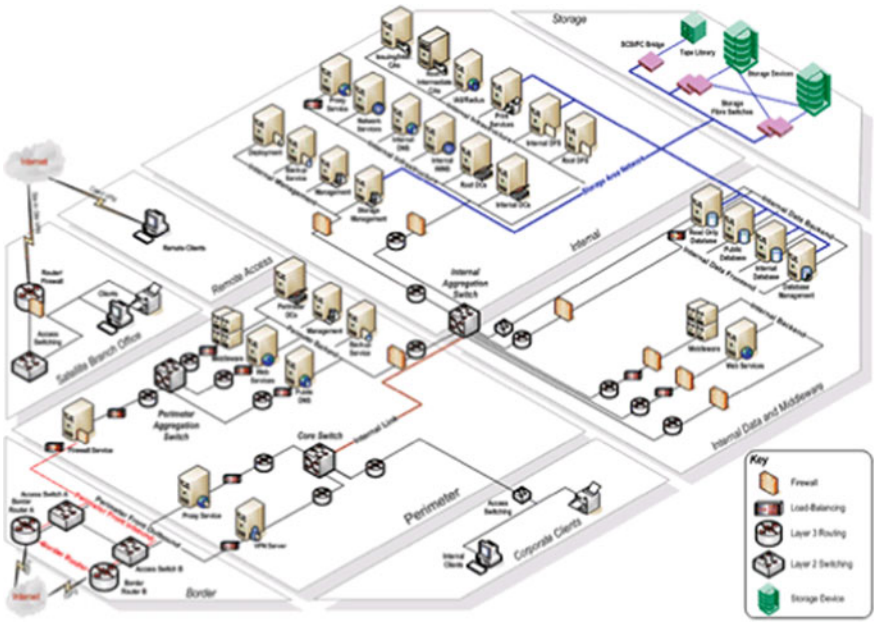



**Fig. 1** IT infrastructure of the data center

The scenario for IT infrastructure building is as follows: architectures of IT infrastructures define the set of services. IT services are provided to three groups of customers. IT services and customers are interrelated within 5 implementation scenarios. The integration of IT services is determined by respective architecture.

## 3 Definition of IT Infrastructure Elements

IT Services shall mean a set of operations aimed at maintaining the following elements in a technically sound state (see Fig. 2):

$$Ss = F(Ss_1,\ Ss_2,\ Ss_3,\ Ss_4,\ Ss_5),$$

where $Ss_1$—network, $Ss_2$—data management, $Ss_3$—IT infrastructure management, $Ss_4$—applications infrastructure, $Ss_5$—security.

Network services, $Ss_1$, include:

$$Ss_1 = F(Sls_{11},\ Sls_{12}),$$

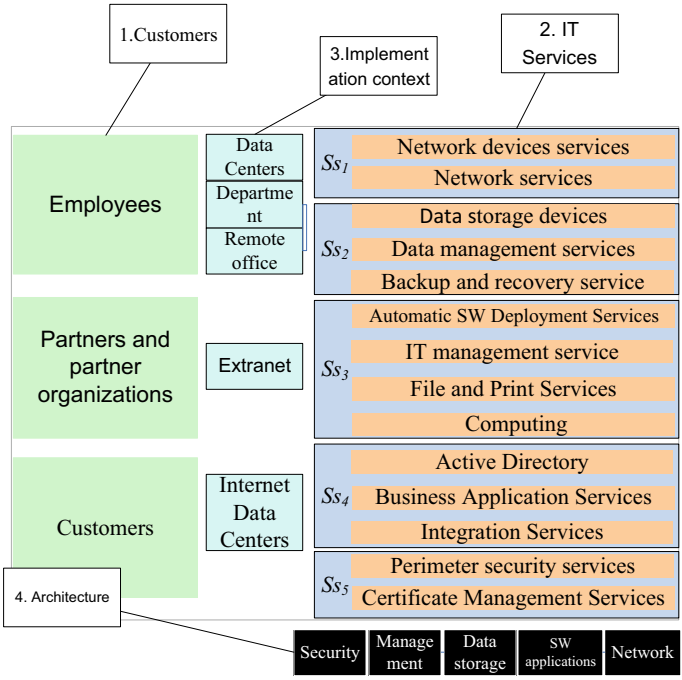



**Fig. 2** IT infrastructure model

where $Sls_{11}$—network devices service, $Sls_{12}$—network service (DNS, DHCP, WINS).

Data management services include:

$$Ss_2 = F(Sls_{21}, Sls_{22}, Sls_{23}),$$

where $Sls_{21}$—data storage device service (DAS, NAS, SAN), $Sls_{22}$—data management service, $Sls_{23}$—backup and recovery service (backup hardware and software, recovery process).

IT infrastructure management services include:

$$Ss_3 = F(Sls_{31}, Sls_{32}, Sls_{33}, Sls_{34}),$$

where $Sls_{31}$—automatic software deployment service, $Sls_{32}$—IT management service, $Sls_{33}$—file and print service, $Sls_{34}$—computing.

Applications infrastructure services include:

$$Ss_4 = F(Sls_{41}, Sls_{42}, Sls_{43}),$$

where $Sls_{41}$—active directory, $Sls_{42}$—online business applications service (ERP, CRM, Exchange, content management services), $Sls_{43}$—integration service.

Security services include:

$$Ss_5 = F(Sls_{51}, Sls_{52}),$$

where $Sls_{51}$—perimeter security service (perimeter PE and internal, proxy/cache services), $Sls_{52}$—certificate management service (PKI).

Business application services, $Sls_{42}$, are detailed and further subdivided into electronic mail service, terminal access service, system support management service, CRM service, etc.

All Enterprise customers, $K$, are divided into three main categories. Where appropriate, the customers may be further grouped within each category:

$$K = F(K_1, K_2, K_3),$$

where $K_1$—Employees, $K_2$—Partners and partner organizations, $K_3$—Customers.

Implementation scenarios, $Sc$ (Fig. 3):

$$Sc = F(Sc_1, Sc_2, Sc_3, Sc_4, Sc_5),$$

where $Sc_1$—Data Centers, $Sc_2$—Department, $Sc_3$—Remote office (telecommunications center, production unit, etc.), $Sc_4$—Internet Data Centers.

Based on the modern structure of the corporation, areas are identified in which Data Centers are consolidated. Data Centers are a set of software and hardware tools that are created in accordance with the described architecture, provide services

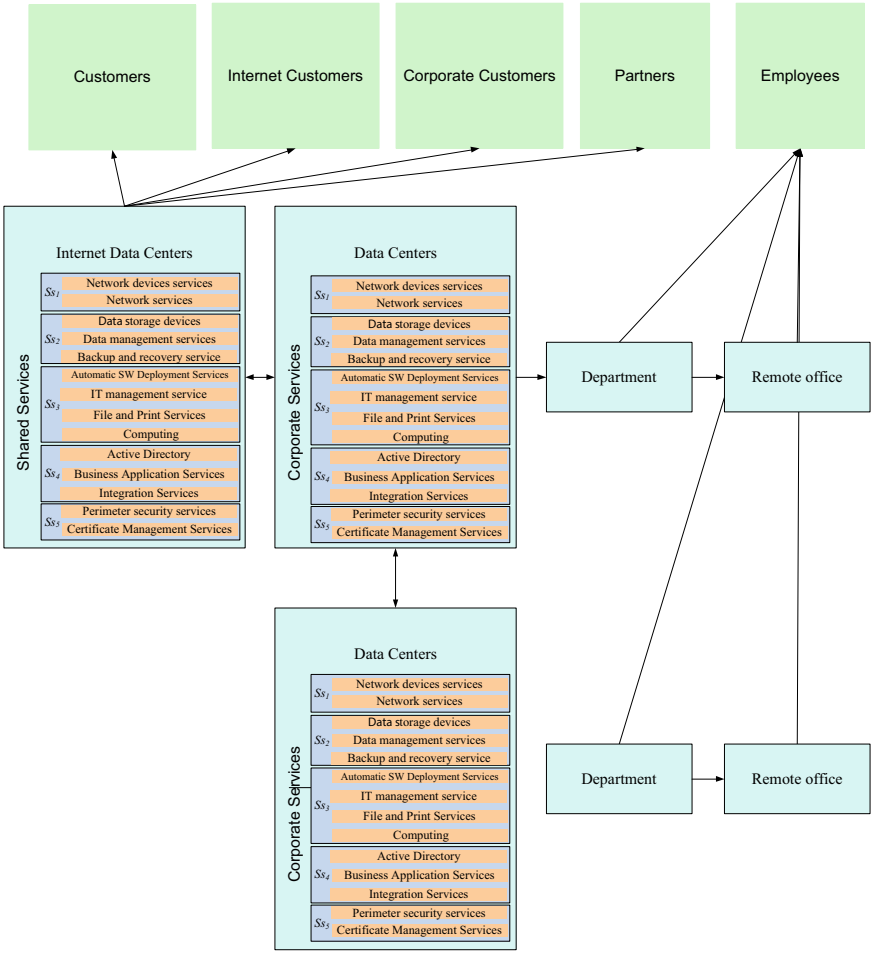



**Fig. 3** IT infrastructure implementation scenarios

exclusively for corporate employees and are serviced by specialized groups of personnel.

The category of departments includes all divisions of the corporation that will not have deployed IT services on their territory. Departments are subdivisions of a corporation that have workstations and user workspaces on their territory, and receive all IT services directly from Data Centers. In their territory, they do not have a any server system, a system that provides a service, or a system that is maintained by specialized groups of personnel (see Fig. 3).

Remote offices have IT service components on their territory, for example, a stand-alone workstation management server. In case of insufficient network bandwidth, some services are decentralized and their modules are moved closer to the clients.

In this case, the remote division of the corporation, which has users on its territory, consumes most of the IT services from the Data Center and has servers that provide a limited set of IT services locally, belongs to the category of remote offices.

Internet Data Centers are deployed at two sites. Internet Data Centers provide shared services (including Internet products) for external clients (Internet and corporate networks), partners, and corporate employees. Internet Data Centers are integrated with Data Processing Centers, however, they are serviced by separate specialized groups of personnel.

Architectures, $A$:

$$A = F(A_1,\ A_2,\ A_3,\ A_4,\ A_5),$$

where $A_1$—Security, $A_2$—Management, $A_3$—Data storage, $A_4$—Software applications, $A_5$—Network.

The scenario for IT infrastructure building is as follows: architectures ($A_i$, $i = 1, 2, \ldots, 5$) of IT infrastructures define the set of services ($Ss_j$, $j = 1, 2, \ldots, 5$). IT services are provided to three groups of customers ($K_n$, $n = 1, 2, 3$). IT services and customers are interrelated within 5 implementation scenarios ($Sc_m$, $m = 1, 2, \ldots, 5$). The integration of IT services is determined by respective architecture $A_i$.

## 4 Mathematical Model of IT Infrastructure

Therefore, we have determined the following principle for building the IT infrastructure: architectures define a set of services, IT services are provided to three groups of customers, IT services and customers are interrelated within five implementation scenarios, five architectures define the integration of IT services.

For the mathematical modeling of the process, we will use the method of optimization of Data Center services provision to selected categories of customers [19].

This task may be solved using the graph theory. As seen from Fig. 4, the abovementioned graph G consists of 4 blocks:

The first block, which characterizes the customer, includes the vertices $v_{11}, \ldots, v_{1n_1}$, which total number is $n_1$ (Fig. 4 n1 = 3), oriented edges beginning in vertices of block $v_{1i}$ ($i = 1, \ldots n_1$) and ending in some vertices of block $v_{2j}$ ($j = 1, \ldots, n_2$); the number of edges is not more than $n_1 * n_2$.

The second block, which characterizes the implementation scenario, includes the vertices, $v_{21}, \ldots, v_{2n_2}$, which total number is $n_2 (n_2 = 5)$, the oriented edges beginning in one of the vertices of block $v_{2j}$ ($j = 1, \ldots, n_2$), and ending in each vertex of block $v_{3k}$ ($k = 1, \ldots n_3$), these edges are available for each vertex of block $v_{2j}$, i.e. the number of oriented edges for this block $= n_2 * n_3$.

The third block, which characterizes the services, includes the vertices $v_{31}, \ldots, v_{3n_3}$ which total number is $n_3 (n_3 = 14)$.

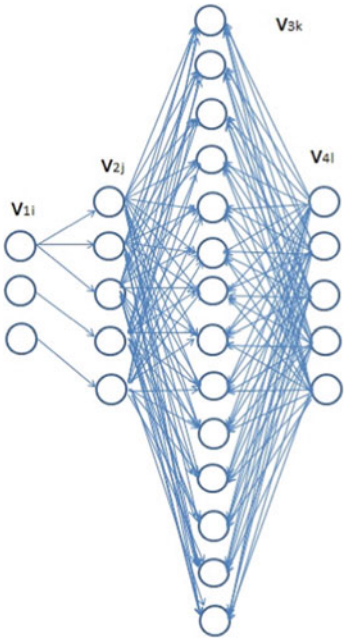



**Fig. 4** Graph G

The fourth block, which characterizes the architectures, includes the vertices $v_{41}, \ldots, v_{4n_4}$, which total number is $n_4(n_4 = 5)$, the oriented edges beginning in one of the vertices of block $v_{4l}$ $(l = 1, \ldots, n_4)$ and ending in each vertex of block $v_{3k}$ $(k = 1, \ldots, n_3)$, these edges are available for each vertex of block $n_4$, i.e. the number of oriented edges for this block $= n_4 * n_3$.

The compatibility matrix for oriented graph G is shown below (Matrix1):

| | $v_{11}$ | ... | $v_{1n_1}$ | $v_{21}$ | ... | $v_{2n_2}$ | $v_{31}$ | ... | $v_{3n_3}$ | $v_{41}$ | ... | $v_{4n_4}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $v_{11}$ | 0 | 0 | 0 | 1 | ... | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| ... | 0 | 0 | 0 | ... | ... | ... | 0 | 0 | 0 | 0 | 0 | 0 |
| $v_{1n_1}$ | 0 | 0 | 0 | 1 | ... | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $v_{21}$ | 0 | ... | 0 | 0 | ... | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | 1 | 1 | 1 | ... | ... | ... |
| $v_{2n_2}$ | 0 | ... | 0 | 0 | ... | 0 | 1 | 1 | 1 | 0 | ... | 0 |
| $v_{31}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $v_{3n_3}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $v_{41}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| ... | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| $v_{4n_4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |

It means that the compatibility matrix for graph G consists of the following blocks (Matrix 2):

$$\begin{pmatrix} 0 & V12 & 0 & 0 \\ 0 & 0 & V23 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & V43 & 0 \end{pmatrix}$$

where *V12* is a binary $n_1 * n_2$ matrix, *V23* and *V43*—unit matrices (all elements = 1). Dimension of *V23* matrix is $n_2 * n_3$, and of *V43* matrix is $n_4 * n_3$.

The incidence matrix of oriented graph G is shown in Matrix 3:

| | $e_{11}$ | … | $e_{1n_1 \times n_2}$ | $e_{21}$ | … | $e_{2n_2 \times n_3}$ | $e_{41}$ | … | $e_{4n_4 \times n_4}$ |
|---|---|---|---|---|---|---|---|---|---|
| $v_{11}$ | 1 | 0 | 0 | 0 | … | 0 | 0 | 0 | 0 |
| … | 0 | 0 | 0 | … | … | … | 0 | 0 | 0 |
| $v_{1n_1}$ | 0 | 0 | 1 | 0 | … | 0 | 0 | 0 | 0 |
| $v_{21}$ | −1 | … | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| … | … | … | … | 1 | 1 | 1 | … | … | … |
| $v_{2n_2}$ | 0 | … | −1 | 1 | 1 | 1 | 0 | … | 0 |
| $v_{31}$ | 0 | 0 | 0 | −1 | −1 | −1 | −1 | −1 | −1 |
| … | 0 | 0 | 0 | −1 | −1 | −1 | −1 | −1 | −1 |
| $v_{3n_3}$ | 0 | 0 | 0 | −1 | −1 | −1 | −1 | −1 | −1 |
| $v_{41}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| … | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| $v_{4n_4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

The number of rows of the incidence matrix $= n_1 + n_2 + n_3 + n_4$.

The number of columns is not more than $n_1 * n_2 + n_2 * n_3 + n_4 * n_3$.

That is, the incidence matrix for the graph G consists of the following blocks (Matrix 4):

$$\begin{pmatrix} E11 & 0 & 0 \\ E21 & E22 & 0 \\ 0 & E32 & E33 \\ 0 & 0 & E43 \end{pmatrix}$$

where *E11* is a binary matrix of maximum dimension $n_1 * (n_1 \times n_2)$, *E21* is a matrix of maximum dimension $n_2 * (n_1 \times n_2)$, which elements are either 1 or 0 (elements of *E11* and *E21* matrices depend on the existence of arcs in the first block), *E22* and *E43* are the matrices of $n_2 * (n_2 \times n_3)$ and $n_4 * (n_4 \times n_4)$ units, respectively, and E*32* and E*33* are the matrices with dimension of $n_3 * (n_2 \times n_3)$ and $n_4 * (n_3 \times n_4)$, respectively, all elements of which $= -1$.

## 5 Choice of Optimal Network Architecture in IT Infrastructure Design

For example, consider the choice of the optimal network architecture. Consider the $A_1$ mesh architecture as a set of parameters:

$$A_1 = F_1(V_1,\ D_1,\ B_1, M_1, K_1,\ P_1,\ Pt_1,\ W_1,\ It_1),$$

where:

1. Relationships between $V_1$ architectures. The purpose of the network architecture is to provide a reliable, scalable, and accessible connection to the network at the physical and logical levels under the requirements of the Enterprise. To ensure that applications have the appropriate level of network service, the network architecture must be designed with the security system architecture in mind, setting certain requirements at the physical (device) and logical (configuration) levels.

   In some cases, the network architecture may depend on the management system architecture. For example, if the management system requires the allocation of a certain network for data transmission for control. The following requirements should be met when designing the network architecture:In some cases, the network architecture may depend on the management system architecture. For example, if the management system requires the allocation of a certain network for data transmission for control. The following requirements should be met when designing the network architecture:
2. Availability $D_1$. The level of network availability is determined by the requirements of the applications that exploit such network. It is impossible and economically impractical to ensure 100% network availability. It is best to determine the availability level of each network device based on the requirements of the applications to be served.

Usually, the redundancy which maintains a high level of network availability is achieved through the use of the following elements:

- $D_{11}$—surplus components. The devices can be designed to achieve redundancy by duplicating their internal components.
- $D_{12}$— "active-active" or "active–passive" clustering. When clustering mechanisms are used, two network devices of each type can be implemented in the architecture to maintain the high level of network availability.
- $D_{13}$—devices or servers of "hot" reserve. You can deploy redundant routers, switches, and firewalls on your network. Therefore, there will be no device on the network that can cause a global outage. In case of a firewall failure, a backup firewall will be activated manually. If the switch stops functioning (it is used in conjunction with a group of network adapters, as described below), the other switch will take full load until the first one is repaired or replaced.
- $D_{14}$—groups of network adapters on hosts. Using a group of network adapters on one host, two network ports are created. Each port is physically connected to a

specific switch, and the switches use a backup protocol (the protocol depends on the vendor equipment used). In addition, the host network adapter driver provides port management as a single logical device. When one port on the network adapter fails, the host continues to communicate through a functioning port, and if the switch stops functioning, all hosts continue to communicate through the network adapter connected to the functioning switch.

When defining network architecture requirements, it is important to analyze all devices located between the customer and the application he or she needs. All devices between the two endpoints must have the same level of availability as the software. Otherwise, the level of availability of the application cannot be considered appropriate. However, if we look at the average uptime of each device in the communication channel, it may turn out that the appropriate levels of availability can be achieved even without the redundancy of some devices.

When calculating the required levels of availability at the Enterprise as a whole, it is necessary to analyze the estimated downtime associated with maintenance for each component, as well as the average uptime. This is necessary to determine the mechanisms (if any), which may be used to achieve high level of availability on the network. It is also important to make sure that the availability requirements of the approved service level agreements can be met.

3. Security $B_1$. Security and performance requirements must be taken into account when designing a network. Employees of the Enterprise responsible for security should determine the security zones that should be observed when designing the network:

- $B_{11}$—multihoming. Servers of the Enterprise can be equipped with several network adapters. This is done, in particular, to split the data streams routed to the Internet customers from the internal data streams, as well as to maintain the appropriate level of performance (two data streams may overload one network adapter, while two adapters manage them well) or to ensure the security. Multihoming of servers involves the use of multiple network adapters or multiple IP addresses on a single server (in the latter case, the server may have one or more adapters). This paragraph will discuss the use of multiple network adapters only (which, in turn, involves the use of many IP addresses).
- $B_{12}$—use of multihoming for security purposes. The advantages and disadvantages of multihoming compared to single-homing are discussed in detail below.

Option 1. Multihoming.
Advantages:

– Improved performance. Multihoming allows to split data streams, in particular, isolate management data flows (generated, for example, by backup or retrieval and remote administration functions) and network adapter used to connect with the Internet.

– Security at the level of individual segments. Multihoming enables to associate each interface with a specific network segment and perform filtering on each interface of TCP/IP ports.

Disadvantages:

– Increased complexity. Multihoming complicates the network design because it requires additional cables, interfaces, inputs, and ports for routing that need to be configured and maintained.
– Additional costs. Multihoming increases expenses for equipment since it requires additional network adapters, cables, and switch ports.

Option 2. Single-homing.
Advantages:

– Lower costs. Single-homing is cheaper than multihoming because it requires fewer network adapters, switch ports, and cables.
– Easier management process. Single-homing is easier to manage than multihoming because it requires fewer network adapters, routing ports, switching ports, and wires.

Disadvantages:

– Single-homing does not provide as much flexibility as multihoming since the network administrator cannot isolate data streams on a network adapter to achieve the appropriate level of performance or security.

- $B_{13}$—security restrictions. Since the network equipment controls the transmission of information within an Enterprise and to the Internet, it must support functions designed to implement security restrictions. The devices that form the network should provide the following capabilities:

– at least authentication and user password verification for remote administrators;
– the best way—encryption of data transmitted for administration and monitoring.

- $B_{14}$—access control lists. You can use Access Control Lists (ACLs) for hosts or network segments. Applying them to network devices allows to control the data which can be transmitted to certain local and virtual local area networks.
- $B_{15}$—service accounts. When servers are used to provide network functionality, all service accounts should be local, not domain. The names and passwords in the accounts must comply with the password assignment guidelines set out in the Enterprise's Security Policy.
- $B_{16}$—network authentication. Network authentication is used to verify the identity of the users trying to connect to the network. This function is typically used in networks where:

– it is difficult to control physical access to network points;
– customers receive remote access through remote access services (for example, in virtual private networks);
– wireless communication is used.

Authentication tools include smart cards, biometrics, IPSec, and Internet Authentication Server (IAS) and in case of wireless communication—802 authentication. More information about these tools and technologies can be found on the websites.

- $B_{17}$—network data encryption. Encryption of data streams in the network protects against malicious data interception and decoding. Common encryption methods are the DES (Data Encryption Standard) and 3DES standards included in the IPSec package, as well as the MPPE (Microsoft Point-to-Point Encryption) standard for PPTR (Point-to-point Tunneling Protocol).

4. Scalability $M_1$. When implementing a network, you need to use intelligent network devices so that network administrators can scale it in the event of increased bandwidth requirements for Data Centers and networks. Sophisticated network devices can perform intelligent packet routing and filtering, resulting in efficient packet movement, often at almost wire speed. Network devices and servers must be able to support port speeds from 10 Mbps to 100 Gbps, as determined by the bandwidth requirements of the environment. Modular switches can be used to increase the number of ports for connecting devices and servers to be added to the environment.
5. Controllability $K_1$. The importance of manageability issues caused the strengthening and deepening links between business needs and network operations. The main tasks of network management include:

   – improving the quality of services;
   – reducing the cost of ownership;
   – reducing security threats.

Managing an Enterprise's environment requires well-built, flexible processes to address business challenges. Environmental management includes administration, problem-solving, and preventive infrastructure development to minimize the number of issues. Management also involves defining service level agreements and verifying the proper quality of services.

- $K_{11}$—network management services. The main tools used by the management services are server management tools that allow to determine network performance, collect event data, compile reports and distribute messages on the network.

  There are many different products designed to perform these functions. To ensure centralized monitoring and control of the services, network and server management tools need to be integrated into the monitoring and warning services implemented in the management system.
- $K_{12}$—system administration. You must be able to remotely and securely manage each network device and server via the network. The Security Policies implemented at the Enterprise can provide for local administration only. However, it is generally recommended to ensure secure remote administration.

  Network management must be secure. Protection of management consoles includes physical protection of devices, use of complex and long passwords, and

protection of network management routes. Technologies such as RADIUS or Secure Shell (SSH) help solve the issue of secure management. Other measures to address this problem include the introduction of encrypted sessions and the use of an isolated network for management. Isolated control may involve physical separation of de-vice management interfaces from all other common routed data streams on the network and restricting control data streams between network hosts. If shared net-work routes are used for management, we strongly recommend using encryption of management data streams.

Management options vary depending on the devices chosen by the Enterprise. At a certain stage of design, it is necessary to decide on the need to perform standardization according to any of the management protocols. Typically, network devices support the following management protocols:

– Telnet;
– Secure Shell (SSH);
– HTTP (Hypertext Transfer Protocol) or HTTPS (Secure Hypertext Transfer Protocol);
– FTP (File Transfer Protocol) or TFTP (Trivial File Transfer Protocol);
– Syslog;
– SNMP (Simple Network Management Protocol).

- $K_{13}$—integrated or dedicated management networks. As noted, network design also involves the development of an Enterprise's Security Policies. Often, there is an access control mechanism operating between the Internet and internal users. At layer 3 and below, it is usually implemented as a firewall with TCP and/or UDP port filters. Although this approach allows for higher levels of security because it monitors the data streams over the network, it often causes difficulties when performing typical management tasks, such as remote administration, backup, or recovery. Many enterprises create separate, or dedicated, management networks that serve management-related operations. However, for security purposes, such networks require control tools too.

The advantages and disadvantages of dedicated and integrated management networks are outlined below.

Option 1. Dedicated management network. An isolated or dedicated management network has the following advantages and disadvantages.

Advantages:

– A dedicated management network is characterized by clearly defined control boundaries. This is achieved through the use of dedicated interfaces and networks to isolate data streams required for management.

Disadvantages:

– Management costs. The management network itself becomes an additional network to be managed.

– Security threat. Depending on a configuration, a dedicated management network can bypass firewalls and other controls and create risks.

Option 2. Integrated management network. The application of a management network that uses the same equipment as the Enterprise's network has the following advantages and disadvantages.

Advantages:

– A management process in an integrated management network is simpler because it has fewer networks, ports, and routing tables to manage.

Disadvantages:

– Potential network congestion. When performing some management-related functions, significant data streams may be generated on the network. If these streams are transmitted over a local network providing customer access to the servers, there may be delays in customer request processing.
– Additional difficulties in case of interruptions. In case of the core network failure, it is difficult to route the data needed to diagnose and fix the problem.

6. Productivity $P_1$. To design a network with the highest possible performance, the following should be taken into account:
   - $P_{11}$—speed of devices. The speed of a device depends on its function (how fast it can perform packet routing or filtering).
   - $P_{12}$—network speed. The speed of network interfaces and communication devices or server ports (for example, 100 Mbps or 1 Gbps).
   - $P_{13}$—filtration. The type of packet filtering (packet verification above layer 3 of the OSI model) determines the required processor power. The higher the layer at which filtering is performed, the more likely the deterioration in performance. If necessary, additional CPUs must be introduced to restore performance levels.
   - $P_{14}$—encryption. The use of encryption, for example in virtual private networks, leads to reduced performance. If the load caused by encryption is too significant and the performance is lower than required, the devices used for encryption must be provided with additional CPU resources—this will restore the performance.
   - $P_{15}$—number of devices. The delay in the network operation generally increases with the increase in the number of devices therein.
7. Support $Pt_1$.Support is the main aspect of network architecture that is often disregarded. Each device introduced into the network environment causes additional costs, i.e. increases the total cost of acquisition and operation of the network. To reduce the full cost of acquisition and operation, it is necessary to identify and purchase devices which design involves the minimum costs associated with their operation. Support options may be expanded with:
   – remote administration tools;

- centralized remote software and hardware upgrades;
- high level of support for industry standards;
- integration with the Enterprise's management system.

8. Consolidation $W_1$. In an ever-growing enterprise, the number of devices is rapidly increasing. New switches appear on each floor and in each server room; some of them are controlled and others are not, some provide a transfer rate of 100 Mbps, while the others—higher. Many enterprises focus on standardization and consolidation to avoid chaos in the network infrastructure. As a result, this process may end up with the choice of a specific method of consolidation.

   - $W_{11}$—consolidation of identical roles. This type of consolidation involves reducing the total number of devices to introduce fewer more powerful devices that can play a certain role at fewer breakpoints. The switch is a likely candidate for this kind of consolidation. Low-power unmanaged switches can replaced with a more powerful and managed switch. With the advent of powerful processors and high-speed network adapters, it has become possible to consolidate complex network devices such as routers, firewalls, and virtual private network devices (remote access servers).
   - $W_{12}$—consolidation of different roles. The boundary between different network devices or server types is gradually disappearing. For example, many routers and switches can act as firewalls or support VPN services. Consolidating multiple components in such multifunction devices or servers helps reduce the number of devices that need to be managed in the environment, thus reducing overall acquisition and operation costs. Companies can resort to this approach based on these considerations.
   - $W_{13}$—administration. Several people can control the operation of different devices or servers. It is not always possible to separate the functions of control over the operation of multifunctional devices so that it corresponds to the structure of the organization.
   - $W_{14}$—larger front to attack. When the functionality of many "terminal" devices (devices connected directly to the Internet) is combined, the security threat increases because such a resulting device is more vulnerable to attacks of various kinds.
   - $W_{15}$—fewer optimization options. For example, if the functionality of a VPN is combined with the functions of a router, the ability to scale services independently is lost. Virtual private network functionality needs encryption that requires CPU resources, and this can reduce routing performance. If these two functions are performed by separate devices, their performance may be optimized separately.

9. Interoperability $It_1$. Elements of the network architecture must interact with each other and with other infrastructure components. It is necessary to ensure interaction at the following levels:

- $It_{11}$—physical. Network facilities must function with other network equipment. For example, they should fit standard equipment racks. Voltage requirements should also be taken into account.
- $It_{12}$—connection. The network hardware must provide the appropriate level of communication to match other elements of the network. For example, it must support a connection using twisted pair or fiber optic cable.
- $It_{13}$—protocols. Support for both Layer 2 and Layer 3 protocols is required. The interoperability opportunities provided by network device manufacturers are usually quite significant. Almost all manufacturers of enterprise-class routers and switches adhere to industry standards published by the IETF (Internet Engineering Task Force) Working Group. These include configuration information and the Routing Information Protocol (RIP). Many published documents from the RFC series are accepted as standards by hardware and software manufacturers.
- $It_{14}$—control. The hardware and software of the network must interact at the management level so that it can be controlled, set up remotely, and made as economical as possible. Comparing management capabilities of equipment purchased from different vendors and of entire routing infrastructure purchased from a single vendor, the benefits of exploiting small, unified sets of management tools and techniques usually encourage enterprises to use a single vendor when possible.

## 6 Examples of PJSC "Ukrtelecom" IT Infrastructure Implementation

The example of PJSC "Ukrtelecom" data transmission network for Ukraine is given below.

The technology for building a VPN in IP/MPLS (see Fig. 5) is used to form a corporate network (Intranet), interact with partner networks (Extranet), and connect to the Internet.

The Data Centers of the Enterprise have the same locations as the Central and Regional Transit nodes of the Enterprise's data network, and accordingly, will be connected to them.

Networks of remote offices and departments access the Data Center through the Regional nodes of the data network.

Ultimately, the Enterprise's network consists of:

1. Six Data Centers (green in the diagram) and two Internet Data Centers.
2. 200–400 remote offices with a standard set of stand-alone equipment and systems (yellow).
3. 1100–1400 networks include only custom workstations (blue color departments).

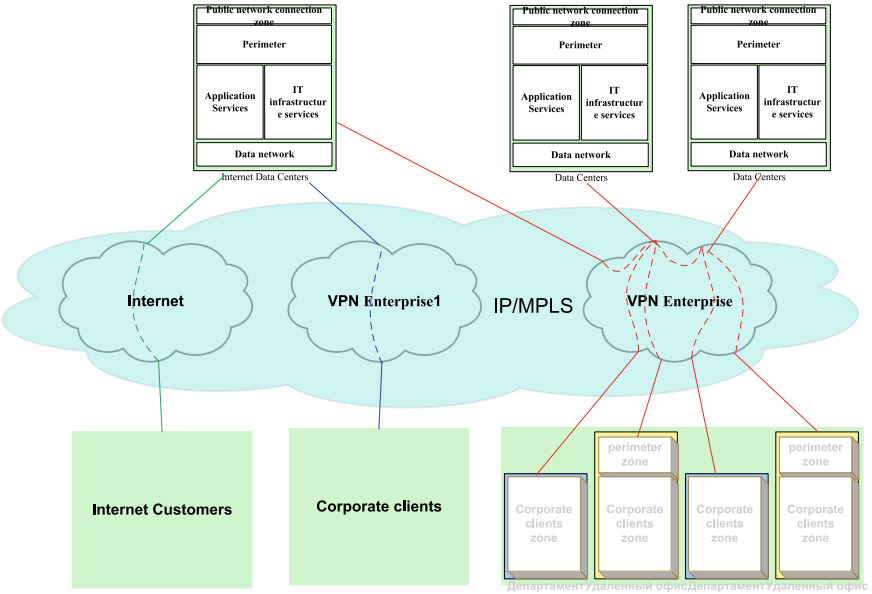



**Fig. 5** Virtual private networks in the enterprise's IP/MPLS network

The Enterprise's network is logically divided into security zones using a perimeter security service. The Enterprise has seven main security zones (see Fig. 6). Data Centers include:

1. Data Network Zone.
2. IT Infrastructure Services Zone.
3. Application Services Zone.
4. Perimeter Zone.

All divisions of the Enterprise, which have employees' workplaces (Management Offices, Information Technologies and Maintenance Centers (ITMC), Production Units, etc.) on their territory and are located within the Enterprise's corporate network include the following zone:

5. Corporate User Zone.

The Data Center connected directly to the Central Node of the Data Transmission Network includes:

6. Public Network Connection Zone.
7. Regional ITMCs are not considered core, and some operational centers that are defined as remote offices include:

   7.1 Perimeter Zone.
   7.2 IT Infrastructure Services Zone.

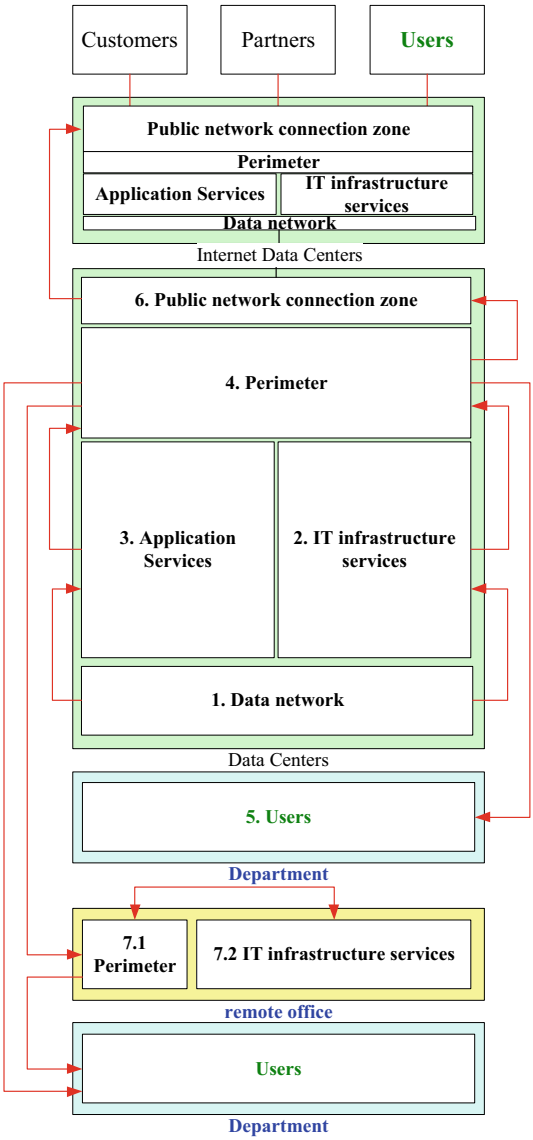



**Fig. 6** Enterprise's network security zones

The proposed model and methodology were first implemented during the construction of data processing service centers in PJSC Ukrtelecom. According to the model, 38,000 work places were transferred to work in data processing service centers using "cloud technology".

A method of optimizing the process of providing certain categories of service for Data Center clients made it possible to build a load matrix of the corporate network of PJSC "Ukrtelecom" and minimize the processing time of customer

requests by the service system, optimize its performance for certain loads for the selected system configuration, which is confirmed by the act of implementation in PJSC "Ukrtelecom".

The proposed model and methodology were used in the development of the feasibility study of the National Telecommunication Network, the need for which is recorded in the Strategic Defense Bulletin of Ukraine.

## 7 Conclusions

1. This paper described the methodological framework for automation of system-wide support. The principles of designing Data Centers as key elements of IT infrastructure have been developed.
2. The paper proposed an improved methodological approach to the design of system IT infrastructure which ensures Data Centers efficiency.
3. We have developed the method to optimize the provision of Data Center services to selected categories of customers, which when integrated determine five architectures (security, management, data storage, software applications, network). The distinctive feature of this method is a clearly defined algorithm of five scenarios for providing access to services, which allows to minimize the time of customer requests processing by the service system and optimize its performance according to a specific load for the selected system configuration.
4. The optimization method of providing certain categories of customers with Data Center services based on the developed standard system architecture of the IT infrastructure was tried and tested.
5. We have formulated the requirements for five Data Center architectures: network, data management, IT infrastructure management, application (software application) infrastructure, and security.
6. This paper discussed the choice of the architecture for the Enterprise Data Centers' IT infrastructure, which is designed to provide a reliable, scalable, and accessible connection to the network at the physical and logical levels according to Enterprise's requirements.
7. To ensure that applications have the appropriate level of network services, the network architecture must be designed with the security system architecture in mind, which sets certain requirements at the structural (device) and logical (configuration) levels.
8. When designing the network architecture, the following requirements were met: accessibility; security; scalability; controllability; support; consolidation; interoperability.
9. In some cases, the network architecture may depend on the architecture of the management system. For example, if the management system requires the allocation of a separate network for the transmission of management data.
10. The model and methodology were implemented in PJSC Ukrtelecom.

## References


1. Redko V, Sergienko I, Stukalo A (1992) Application systems. Architecture, construction, development, Naukova dumka, Kyiv, 320 p
2. Repin V, Elyferov V (2004) Process approach to management: business process modeling. RIA "Standards and quality", Moscow, 408 p
3. Rinks D (1986) A heuristic approach to aggregate production scheduling using linguistic variables: methodology and Application. In: Yager RR (ed) Fuzzy set and possibility theory, recent advances. Radio and Communications, Moscow, p 40
4. Globa L, Dovgiy S, Kopiika O, Kozlov O (2021) Approach to building uniform information platform for the national automated ecological information and analytical system. In: CEUR workshop proceedings, vol 3021, pp 53–65
5. Kopiika O, Skladannyi P (2021) Use of service-oriented information technology to solve problems of sustainable environmental management. In: CEUR workshop proceedings, vol 3021, pp 66–75
6. Barabash O, Kopiika O, Zamrii I, Sobchuk V, Musienko A (2019) Fraktal and differential properties of the inversor of digits of QS-representation of real number. Modern Math Mech Fund Probl Challenges, 79–95
7. Kopeika O, Tereshchenko A (2001) Wind-power transforming systems. J Math Sci 104(6):1631–1634
8. ITIL® V3 Foundation Course Glossary (2021). https://itil.it.utah.edu/downloads/ITILV3_Glossary.pdf
9. What is IT Infrastructure? https://www.ecpi.edu/blog/what-is-it-infrastructure
10. Beginner's Guide to IT Infrastructure Management. Smartsheet (2020). https://www.smartsheet.com/it-infrastructure-management-services-guide
11. What is infrastructure (IT infrastructure)? - Definition from WhatIs.com. SearchDataCenter (2019). https://searchdatacenter.techtarget.com/definition/infrastructure
12. Reference architectures MSA (2005) Microsoft Ukraine. BHN, Kyiv, 352 p
13. Information technology - practical rules for information security management, ISO/IEC 17799. International Standard (2000)
14. Jonathan J (2010) BICSI data center standard: a resource for today's data center operators and designers. BICSI News Mag 28
15. Susan N (2011) Standardization and modularity in data center physical infrastructure, vol. 4. Schneider Electric
16. Telecommunications industry association standards. http://www.tiaonline.org/standards/
17. Telecommunications Infrastructure Standard for Data Center (2005). TIA Standard TIA-942. Telecommunications Industry Association, 135 p
18. ANSI/BICSI 002-2011 (2011) Data center design and implementation best practices, Committee Approval-January 2011 First Published, 367 p
19. Kopiika OV (2014) Methodology for the synthesis of the information and communication systems based on unified information platform. Abstract of the dissertation of the doctor of technical sciences, 41 p

# Comparative Analysis of Object Detection Methods in Computer Vision for Low-Performance Computers Towards Smart Lighting Systems




**Ivan Matveev, Kirill Karpov, Maksim Iushchenko, Dmitrii Dugaev, Ivan Luzianin, Eduard Siemens, and Ingo Chmielewski**

**Abstract** Nowadays the object detection research based on machine learning techniques is focused on improving the accuracy and the detection speed of a given algorithm. However, most of such approaches assume substantial amount of computational resources available to the algorithm to make it fairly efficient. Therefore, a vast majority of them are hardly feasible on low-powered and less-capable embedded IoT devices, where the object detection tasks are equally common and usually even more challenging—considering a huge diversity in environmental conditions, camera positions and resolutions, outdoor illuminations and deployments in both urban and offline remote environments. This paper presents a comparative research of different machine-learning object detection approaches targeted specifically towards

I. Matveev (✉) · K. Karpov (✉) · M. Iushchenko (✉) · I. Luzianin (✉) · E. Siemens (✉) · I. Chmielewski
Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany
e-mail: ivan.matveev@dexor.de

K. Karpov
e-mail: kirill.karpov@hs-anhalt.de

M. Iushchenko
e-mail: maksim.iushchenko@hs-anhalt.de

I. Luzianin
e-mail: ivan.luzianin@hs-anhalt.de

E. Siemens
e-mail: eduard.siemens@hs-anhalt.de

I. Chmielewski
e-mail: ingo.chmielewski@hs-anhalt.de

K. Karpov · M. Iushchenko
Department of Transmission of Discrete Data and Metrology, Siberian State University of Telecommunications and Information Sciences, Kirova Street 86, 630102 Novosibirsk, Russia

D. Dugaev (✉)
The City University of New York—The Graduate Center, 365 5th Avenue, New York 10016, USA
e-mail: ddugaev@gradcenter.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_10

the low-performance embedded computers in a context of smart street lighting applications. The comparison includes real and synthetic datasets preprocessed for a given detection method, conducted on a Raspberry Pi SoC platform.

**Keywords** Object detection · Single-board computer · Low-performance · Machine learning · Deep learning · Raspberry Pi · IoT · Synthetic data

## 1 Introduction

Computer Vision (CV) is a rapidly developing field in the modern data science [22]. It plays a significant role in the concept of a smart city, e.g. in such services as object tracking [19], public security, e-commerce and others. One of such services is a smart street lighting [38] aimed at energy savings via efficient power control of the lamps. The CV enables street lighting to be adaptive to pedestrian and car traffic by detecting and classifying moving objects.

Object Detection Models (ODM) as a basis of CV implementations must meet the strict requirements to accuracy and performance. Since it is hardly possible to cover all the requirements at the same time, modern algorithms tend to find a trade-off. This trade-off may be achieved in different ways, i.e. by reducing the complexity of the model, using different pre-processing methods, etc. In the paper, the quality metrics of Dimensional Based Object Detection (DBOD) algorithm used in smart lighting and proposed in [18] are selected and compared with the metrics from the state-of-the-art object detection methods.

In case of smart lighting, each lamp pole is equipped with a CCTV camera directed towards a street, and a processing unit. The processing unit detects pedestrians, bicyclists and cars from images provided by the camera. After detecting an object, a group of neighboring lamps is switched on in a direction of the object's movement. The units of different lamps are connected through a decentralized network. Since video traffic may create a significant network load while being transmitted to a centralized object detection node, the detection process is assumed to be performed by the units locally and independently. Therefore, the task of object detection in this scenario has a number of challenges:

- The images are grayscale and have poor contrast, since they are captured in near-infrared spectrum.
- It is necessary to make a switching decision in real-time because objects can move with high velocity. The detection speed of single-board IoT devices is slow because their processing units do not include additional instructions existing in "big" non-embedded processors.
- The image resolution is lowered to 320 × 240 px in order to reduce amount of processing data.
- Deficiency of labeled training data from real street environments that can be compensated by the synthetic data [17].

Algorithm's quality is usually estimated by well known performance metrics [22], such as precision, recall, f-measure and mean Average Precision. In this paper, due to the task specificity, the quality metrics additionally include performance measurements of an algorithm. Already existing detection frameworks that can be deployed on Raspberry Pi are considered in the current comparative study.

## 2 Related Work

The object detection methods have been rapidly developing for the last two decades. The modern approaches are able to detect objects with extremely high accuracy. However, for the most methods, the higher precision results in the lower detection speeds. Thus, state-of-the-art algorithms try to find a trade-off between accuracy and performance [21]. Low frame rate is acceptable in a number of applications, while it remains critical for detection of fast moving objects. Therefore, it is challenging to perform real-time detection by complex algorithms, such as neural networks, deployed on the single-board computers.

In Machine Learning (ML) it is common to distinguish classical and deep learning methods. Widely used classical CV object detection methods are Viola-Jones, HOG, DPM, ICF. Among deep learning approaches the most common are SSD- and RCNN-based networks.

### *2.1 Classical Models*

At the beginning of CV era, the computational complexity of the model had a significant impact due to low performance of computers. Classical object detection methods include two-stage ML-based classifiers. At the first stage, different features are hereby extracted from the image. After that, the features are classified by a predefined classifier. The classification quality hereby depends on type of extracted features and used classification approach.

One of the first detection algorithms called Viola-Jones detector (VJ) was originally introduced in 2001 by P. Viola and M. Jones [31, 32]. A sliding window is used to calculate HAAR-like features over an image. HAAR-like features are combined into HAAR cascades, which are classified by ensemble learning methods such as AdaBoost or Random Forests. The VJ was the fastest solution at that time.

Integral Channel Features (ICF) detector proposed by P. Dollar et al. in 2009 [8] with the improvement in [3, 4] presents an extension of a VJ detector. In addition to the HAAR features, ICF is able to use gradients, filtered channel features [37], convolutional channel features [34]. J. Sochman proposed ICF-like solution using Wald's sequential probability ratio test and AdaBoost algorithm for decision-making [28]. R. Juranek et al. adapted ICF detector for video steaming tasks [16, 36]. ICF-based detectors are still improving nowadays [5].

The solution of N. Dalal [6] extracts Histograms of Oriented Gradients (HOG) and use the Support Vector Machine (SVM) for classification. Deformable Part-based Models (DPM) [9] were introduced in 2008 as an extension to HOG. Another HOG extension called Co-Occurrence HOG (CoHOG) has been proposed for face recognition [7], human recognition [33] and changing environments [35].

## 2.2 Deep Learning Models

The development of GPU-based computing contributed to widespread usage of Deep Learning algorithms for object detection. In practice, they always outperform traditional ML-based solutions in terms of detection accuracy, however they require extremely large resources for learning and are usually slower in recognition. Among all existing neural network structures, the Convolutional Neural Networks (CNN) are the most suitable solutions for object detection. A variety of CNN modifications have been developed during last two decades [15].

Like in the classical CV methods, the first CNN-based object detectors also had two stages. At the first stage, the image is divided into Regions of Interests (ROI). At the second stage, the regions are processed by CNN.

The first deep learning solution is called Region-based Convolutional Neural Network (R-CNN) [26]. R-CNN showed high accuracy in comparison to existing alternatives, however it was very slow. Fast R-CNN [10], Faster R-CNN [25] were designed to increase a detection speed of the original R-CNN.

In parallel, You Only Look Once (YOLO) algorithm was developed by R. Jones et al. in 2015 [23] to create a one-step process involving detection and classification. In this algorithm, the bounding boxes and class predictions are made at once in contrast to the other detection algorithms like R-CNN [24]. The multi-label classification is performed by independent logistic classifiers. This approach enables the algorithm to make predictions faster and computationally lighter, in comparison with the two-stage detectors.

Despite the big variety of proposed neural network configurations, it is still challenging to achieve the highest accuracy and detection speed simultaneously. Two stage detectors usually have higher accuracy, while one-stage are typically faster. The trade-off can be found by using the optimal strategy proposed in [29].

## 2.3 Dimensional Based Models

Dimensional Based Models (DBM) classify objects depending on variety of their physical sizes. Physical dimensions of an object can be measured by sensors, such as ultrasonic, radio wave, infrared and their combinations. Dimensions are represented as changes of physical signals in time. Therefore, the DBM classification task can be considered as a pattern recognition in time series. Examples of DBM based on

different signals are given in [1, 2, 20, 30]. Instead of time series, the DBOD algorithm uses a sequence of images as input data. All the DBM methods require preliminary knowledge about sensor installation parameters for object classification.

## 3 Testbed Description

The smart lighting systems are applied in streets with a large variety of environmental conditions like illumination level, traffic distribution, camera position, etc. It is often difficult to train a classification model on the data obtained from several real scenes. Instead, the model has to be trained using a substantial amount of generic urban scenes. However, obtaining the data containing various scenes of a real city is challenging. Full reconstruction of several scenes identically reflecting the real ones in a virtual environment is ineffective, since it restricts training dataset. Therefore, the synthetic dataset represents a generic urban environment under various conditions. To evaluate accuracy and performance of the algorithms being trained on real and synthetic datasets, the following 4 scenarios have been used:

- Real-real—the model is trained and tested using real dataset only.
- Synth-synth—the model is trained and tested using synthetic dataset only.
- Real-synth—the classifier is trained using real data and tested on synthetic images.
- Synth-real—training is performed on synthetic dataset, while real images are used for testing.

A split of 30% from general dataset is used for testing. The Real-synth and Synth-real scenarios are used to prove validity of the synthetic dataset relatively to the real dataset. All the images are grayscale and rescaled to a resolution of 320 × 240 px. The selected models were trained on real and synthesized datasets with the following amount of frames in the training and testing subsets, presented in Table 1.

The images in real dataset are captured in a nighttime environment via a camera with a near-infrared illuminator. The real dataset represents most common urban scenes: parking lots and city streets with varying traffic intensity. Examples of dataset images are shown in Fig. 1.

The synthetic dataset contains scenes in city environment obtained using Unity3D based smart-light urban simulator [17]. The simulated light sources are an infra-red camera spotlight, street lights and vehicle/bicyclist lamps. The camera is set on street lights (height of 3–5 m above the ground level) and directed to reconstructed

**Table 1** Number of images used for training and testing

| | Real | Synthesized |
|---|---|---|
| Training, frames | 2201 | 6621 |
| Testing, frames | 943 | 2837 |
| Total, frames | 3144 | 9458 |



**Fig. 1** Images from the real dataset

**Table 2** Properties comparison of synthetic datasets to publicly available ones

| | Real (open access) | Synthetic |
|---|---|---|
| Nighttime scenes | Very limited | Any |
| Camera parameters | Usually not provided | Any |
| Illumination | Daytime/street lights/IR | Daytime/street lights/IR |
| Camera position | Arbitrary | Smart-lighting specific |

carriageway with sidewalks. Such kind of scene configuration is expected in the scenario of future smart lighting and smart city systems.

The main reason for synthetic dataset generation is lack of such specific kinds of images in public access (Table 2). Moreover, the synthetic generator provides detailed parameters about the scene configuration, which are required by the DBOD algorithm, e.g. cameras' focal length, height, incline, sensor dimensions, etc. Some samples from the generated synthetic dataset are shown in Fig. 2. The per-class number of objects for each dataset is shown in Table 3.

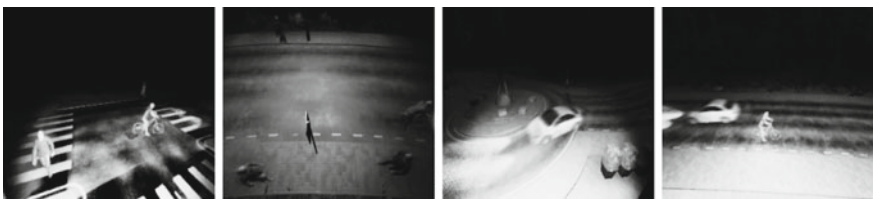


**Fig. 2** Images from the synthetic dataset

**Table 3** Characteristics of the dataset used for evaluating the CNN-based models

| Dataset | Pedestrian | Bicyclist | Vehicle |
|---|---|---|---|
| Real | 2207 | 1232 | 895 |
| Synthetic | 5427 | 3069 | 1070 |

# 4 Comparison of the Models

## *4.1 Deep Learning Models*

The following implementations of deep learning models have been chosen for comparison with the DBOD algorithm:

- YOLOv3 tiny.
- ssdlite_mobilenet_v2.
- ssd_mobilenet_v1.
- faster_rcnn_inception_v2.
- ssd_mobilenet_v2_quantized.

YOLO has a special, so-called "tiny", implementation designed for constrained environments such as single-board computers. Since the task requires real-time image processing on the IoT devices, the tiny version has been tested [13]. The used YOLOv3 tiny model is implemented within a Darknet framework.

Sdlite_mobilenet_v2 is a light modification of SSD (Single-Shot Detection) model, providing slightly better detection performance than the original SSD model. It is based on MobileNetV2 neural network architecture [27].

Ssd_mobilenet_v1 is a classic SSD model based on MobileNetV1 neural network architecture [12]. It provides better accuracy, comparing to SSDLite, however, the object detection speed might be less efficient.

Faster_rcnn_inception_v2 is an implementation of FRCNN model [25], providing even better accuracy comparing to SSD models. However, a certain disadvantage of this model can be in the considerably higher amount of computational resources required by the model during the detection process, comparing to the rest of the models. This drawback might negatively impact its deployment on the embedded devices.

Ssd_mobilenet_v2_quantized is an 8-bit quantized SSD model, which provides worse accuracy but much higher detection speeds, which is important on the embedded platforms. The model was trained using the TensorFlow framework and then was converted to a TFLite (TensorFlow Lite) format, specifically used on low-performance machines [14].

The above 4 deep learning models are the part of the TensorFlow-v1 framework. TensorFlow framework presents an open-source library for machine learning and AI, presenting a flexible set of tools to train and evaluate deep-learning models on given pre-labelled datasets. These machine learning models for object classification were selected from the official TensorFlow 1 Detection Model Zoo for further training and evaluation.

The models have been tested on real and synthesized test datasets. The obtained accuracy metrics for all the testing scenarios are represented in the Table 4. The mean Average Precision (mAP) is calculated at Intersection over Union (IoU) of 0.5.

As it is shown in the Table 4, the CNN-based models were trained well enough to recognize the objects of three classes both on real and synthesized data (scenarios

**Table 4** Experimental quality metrics

| | mAP; IoU = 0.5 | | | |
|---|---|---|---|---|
| | Real-real | Synth-synth | Real-synth | Synth-real |
| DBOD | 0.85 | 0.85 | 0.79 | 0.76 |
| YOLOv3 tiny | 0.97 | 0.96 | 0.03 | 0.07 |
| ssdlite_mobilenet_v2 | 0.93 | 0.89 | 0.10 | 0.02 |
| ssd_mobilenet_v1 | 0.88 | 0.85 | 0.17 | 0.01 |
| faster_rcnn_inception_v2 | 0.98 | 0.95 | 0.10 | 0.36 |
| ssd_mobilenet_v2_quantized | 0.95 | 0.52 | 0.03 | 0.10 |

Real-real and Synth-synth) with high precision. However, these models struggle to be tested in cross scenarios (Real-synth and Synth-real), where they demonstrate very low precision. This result is caused by the difference between real and synthesized datasets in terms of scene parameters: illumination, image contrast, camera parameters, etc.

RCNN-based detector shows slightly better results in the Synth-real scenario (Table 4), however the achieved prediction accuracy is still not appropriate. In addition, this is the slowest and the most complicated model as shown in Table 5.

All the models were deployed on a Raspberry Pi 4 platform to collect some real-time performance metrics for each of the tested model. The performance metrics are associated with how fast a given model handles the object detection, measured in an average number of processed frames per second (FPS). Additional metrics included the consumed CPU and memory resources during the operation. CPU usage is related to the computational complexity of the model, memory usage depends on the framework optimization and average FPS shows the performance of the algorithm on the platform.

**Table 5** Average performance metrics of the models deployed on Raspberry Pi 4

| Model name | FPS | CPU, % | Memory, MB |
|---|---|---|---|
| DBOD | 128.00 | 247 | 112 |
| YOLOv3 tiny | 0.46 | 373 | 122 |
| ssdlite_mobilenet_v2 | 2.34 | 247 | 380 |
| ssd_mobilenet_v1 | 2.21 | 257 | 395 |
| faster_rcnn_inception_v2 | 0.19 | 377 | 767 |
| ssd_mobilenet_v2_quantized | 3.47 | 287 | 197 |
| CPU and memory usage have been measured using top utility | | | |

### 4.2 DBOD Algorithm

The DBOD algorithm operation can be divided into the following stages:

- Pre-processing, Background Subtraction (BS) and filtering—a sequence of contrast enhanced images is processed by a BS method to extract Regions of Interest (RoI). Morphological filtering is applied to a resulting foreground mask to remove undesired RoIs.
- Feature extraction and classification—at this stage geometrical parameters of an object in image are transformed into features describing the object in a real-world coordinate system. Such a transformation is performed using known camera extrinsic and intrinsic parameters. The resulting features are estimated object height ($m$), width ($m$), contour area ($m^2$), distance ($m$), camera height ($m$) and tilt angles ($deg$). These features are used for classification by computationally inexpensive ML method, for instance, by a logistic regression classifier. For training a classifier, the number of features was increased by the second degree polynomial combinations.

The DBOD algorithm uses the same dataset as other testing models, however, it has some differences due to functional peculiarities of the algorithm. The method uses an additional class noise as a negative class. The noises are the RoIs appearing due to rapid changes in scene illumination or oscillation of background objects (e.g. trees). The class noise is not the target class, but the supporting one.

Other deviation of the dataset from the datasets described in Sect. 3 is that the vehicle lights reflections from scene surfaces are considered as part of the vehicle object and, therefore, are labeled as vehicle class. Static objects are excluded from the dataset, since the DBOD is an algorithm based on background subtraction that make it possible to segment only objects which change location between successive frames. Considering the above factors, characteristics of the dataset (Table 6) for this method are slightly different from the datasets for the other methods.

As can be seen from the Tables 4 and 5, the DBOD algorithm is the fastest solution which can operate adequately in all the testing scenarios. Therefore, the deeper study of the DBOD algorithm has been carried out. The detailed per-class accuracy metrics are represented in the Table 7.

In contrast to the deep learning models which consider object appearance in image, the algorithm shows appropriate accuracy in cross testing (Real-synth, Synth-real), since the detection is performed based on rough geometrical measures of an object. Nevertheless, the accuracy in cross testing is lower than in direct scenarios due

**Table 6** Number of objects used for evaluating the DBOD algorithm

| Dataset | Noise | Pedestrian | Bicyclist | Vehicle |
|---|---|---|---|---|
| Real | 26,566 | 2086 | 1131 | 575 |
| Synthetic | 47,138 | 1783 | 1119 | 1793 |

**Table 7** Accuracy metrics for the DBOD algorithm

| Class | Metric | Real-real | Synth-synth | Real-synth | Synth-real |
|---|---|---|---|---|---|
| Noise | f1-score | 0.98 | 0.97 | 0.95 | 0.98 |
| | Precision | 0.99 | 0.97 | 0.94 | 0.98 |
| | Recall | 0.98 | 0.96 | 0.96 | 0.97 |
| | AP | 0.99 | 0.99 | 0.99 | 0.99 |
| Pedestrian | f1-score | 0.84 | 0.79 | 0.73 | 0.78 |
| | Precision | 0.79 | 0.76 | 0.75 | 0.71 |
| | Recall | 0.91 | 0.83 | 0.72 | 0.87 |
| | AP | 0.87 | 0.87 | 0.76 | 0.79 |
| Bicyclist | f1-score | 0.76 | 0.62 | 0.62 | 0.56 |
| | Precision | 0.75 | 0.62 | 0.54 | 0.72 |
| | Recall | 0.77 | 0.62 | 0.73 | 0.46 |
| | AP | 0.71 | 0.60 | 0.60 | 0.60 |
| Vehicle | f1-score | 0.80 | 0.88 | 0.59 | 0.70 |
| | Precision | 0.82 | 0.86 | 0.89 | 0.61 |
| | Recall | 0.78 | 0.91 | 0.44 | 0.81 |
| | AP | 0.84 | 0.93 | 0.79 | 0.66 |
| All | mAP macro | 0.85 | 0.85 | 0.79 | 0.76 |
| | f1-score macro | 0.85 | 0.82 | 0.72 | 0.75 |

to imperfections of the proposed simulation. When testing the algorithm in direct scenarios (Real-real or Synth-synth), similar average accuracy metrics were obtained.

The classification result is summarized in the corresponding confusion matrices (Fig. 3). In both test scenarios there is a problem of misclassification between the pedestrian and cyclist classes. This problem can be explained by the similar geometric parameters of these objects in the image plane at some angles of movement. The class vehicle has the lowest false positive (FP) error rate relative to any other class of the target group due to the rare occurrence of special 2D projection cases, where the features of vehicle and any other classes are intersecting. However, in some projections, vehicle can be interpreted as noise, that leads to high values of the false negative (FN) coefficient (0.19, Fig. 3a).

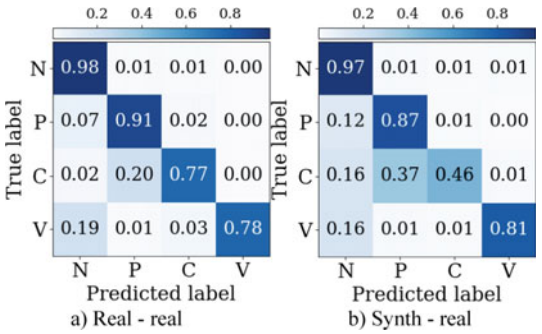



**Fig. 3** Multi-class confusion matrices

## 5 Conclusion

In the course of the study, a comparative analysis of the state-of-the-art deep learning models and the DBOD algorithm in terms of accuracy and performance on the nighttime datasets was carried out.

In the testing Synth-synth and Real-real scenarios, all the deep learning solutions provide a more accurate detection ($mAP > 95$) than DBOD. The DBOD algorithm has the lowest accuracy in these scenarios ($mAP = 0.85$) compared to the CNNs, which is still applicable for the target scenarios.

In contrast to the DBOD, in the cross-testing scenarios (Real-synth and Synth-real), the deep learning models show extremely low accuracy. Such results are explained by the peculiarities of synthetic and real datasets (scenes have different physical parameters), as described in the Sect. 3. In most cases, it turned out that the training and test datasets have no common features, which does not allow training neural networks on the proposed synthetic data. The DBOD algorithm is more much robust in these scenarios that is confirmed by detection accuracy of $mAP = 75$ (Synth-real) and $mAP = 79$ (Real-synth).

The one-stage detectors show the highest detection speed among the CNNs ($0.46 <= FPS_{SSD} <= 2.34$) on Raspberry Pi 4. Being a two-stage detector, Faster RCNN is slower ($FPS = 0.2$). Thus, the use of CNN-based methods is hardly possible for detecting fast moving objects such as cars on the considered embedded computing platforms. The DBOD algorithm has the highest detection speed among the considered ones ($FPS = 128$).

Some studies show [11] that neural networks can be trained on synthetic data and have a higher accuracy than DBOD, when the real environment is reconstructed identically. Nevertheless, with the absence of sufficient amount of real data and the lack of accurate information about the environment characteristics, training the DBOD algorithm on generic synthetic data provides quite high precision, while neural networks are ineffective in these conditions. It makes the DBOD algorithm the most suitable approach for the smart lighting systems.

## References


1. An J, Choi B, Kim H, Kim E (2019) A new contour-based approach to moving object detection and tracking using a low-end three-dimensional laser scanner. IEEE Trans Veh Technol 68(8):7392–7405
2. Bartsch A, Fitzek F, Rasshofer RH (2012) Pedestrian recognition using automotive radar sensors. In: Advances in radio science, vol 10. Copernicus GmbH, pp 45–55. https://doi.org/10.5194/ars-10-45-2012
3. Benenson R, Mathias M, Timofte R, Van Gool L (2012) Pedestrian detection at 100 frames per second. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 2903–2910
4. Benenson R, Mathias M, Tuytelaars T, Van Gool L (2013) Seeking the strongest rigid detector. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3666–3673
5. Chesalin A, Grodzenskiy SY (2019) Modification of the waldboost algorithm to improve the efficiency of solving problems of technical diagnostics of electrical systems. In: 2019 international conference on electrotechnical complexes and systems (ICOECS). IEEE, pp 1–4
6. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR 2005), vol 1. IEEE, pp 886–893
7. Do TT, Kijak E (2012) Face recognition using co-occurrence histograms of oriented gradients. In: 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1301–1304
8. Dollár P, Tu Z, Perona P, Belongie S (2009) Integral channel features. In: Proceedings of the British machine vision conference, CVPR 2001. BMVC Press, pp 91.1–91.11
9. Felzenszwalb P, McAllester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
10. Girshick R (2015) Fast R-CNN. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 1440–1448
11. Hattori H, Naresh Boddeti V, Kitani KM, Kanade T (2015) Learning scene-specific pedestrian detectors without real data. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3819–3827
12. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861 [cs]
13. Huang R, Pedoeem J, Chen C (2018) Yolo-lite: a real-time object detection algorithm optimized for non-gpu computers. In: 2018 IEEE international conference on big data (big data). IEEE, pp 2503–2510
14. Jacob B, Kligys S, Chen B, Zhu M, Tang M, Howard A, Adam H, Kalenichenko D (2018) Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2704–2713
15. Jiao L, Zhang F, Liu F, Yang S, Li L, Feng Z, Qu R (2019) A survey of deep learning-based object detection. IEEE Access 7:128837–128868
16. Juranek R (2008) Detection of dogs in video using statistical classifiers. In: International conference on computer vision and graphics. Springer, pp 249–259
17. Karpov K, Luzianin I, Iushchenko M, Siemens E (2021) Urban environment simulator for train data generation toward cv object recognition. In: Proceedings of international conference on applied innovations in IT (ICAIIT)
18. Matveev I, Karpov K, Chmielewski I, Siemens E, Yurchenko A (2020) Fast object detection using dimensional based features for public street environments. Smart Cities 3(1):93–111
19. Matveev I, Karpov K, Siemens E, Yurchenko A (2020) The object tracking algorithm using dimensional based detection for public street environment. Eur Phys Tech J 17(2):123–127

20. Noyce DA, Gajendran A, Dharmaraju R (2006) Development of bicycle and pedestrian detection and classification algorithm for active-infrared overhead vehicle imaging sensors. Transp Res Rec 1982(1):202–209. https://doi.org/10.1177/0361198106198200125
21. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, Riordan D, Walsh J (2019) Deep learning vs. traditional computer vision. In: Science and information conference. Springer, pp 128–144
22. Padilla R, Netto SL, da Silva EA (2020) A survey on performance metrics for object detection algorithms. In: 2020 International conference on systems, signals and image processing (IWSSIP). IEEE, pp 237–242
23. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, Las Vegas, pp 779–788. https://doi.org/10.1109/CVPR.2016.91
24. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
25. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, vol 28. Curran Associates, Inc.
26. van de Sande KEA, Uijlings JRR, Gevers T, Smeulders AWM (2011) Segmentation as selective search for object recognition. In: 2011 international conference on computer vision, pp 1879–1886. https://doi.org/10.1109/ICCV.2011.6126456
27. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 4510–4520
28. Sochman J, Matas J (2005) Waldboost-learning for time constrained sequential detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR 2005), vol. 2. IEEE, pp 150–156
29. Soviany P, Ionescu RT (2018) Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction. In: 2018 20th international symposium on symbolic and numeric algorithms for scientific computing (SYNASC). IEEE, pp 209–214
30. Stiawan R, Kusumadjati A, Aminah NS, Djamal M, Viridi S (2019) An ultrasonic sensor system for vehicle detection application. J Phys Conf Ser 1204:12–17. https://doi.org/10.1088/1742-6596/1204/1/012017
31. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, CVPR 2001, vol 1. IEEE, p I
32. Viola P, Jones M et al (2001) Robust real-time object detection. Int J Comput Vision 4(34–47):4
33. Watanabe T, Ito S, Yokoi K (2010) Co-occurrence histograms of oriented gradients for human detection. IPSJ Trans Comput Vis Appl 2:39–47
34. Yang B, Yan J, Lei Z, Li SZ (2015) Convolutional channel features. In: Proceedings of the IEEE international conference on computer vision, pp 82–90
35. Zaffar M, Ehsan S, Milford M, McDonald-Maier K (2020) Cohog: a light-weight, compute-efficient, and training-free visual place recognition technique for changing environments. IEEE Robot Autom Lett 5(2):1835–1842
36. Zemcik P, Juranek R, Musil P, Musil M, Hradis M (2013) High performance architecture for object detection in streamed videos. In: 2013 23rd international conference on field programmable logic and applications. IEEE, pp 1–4
37. Zhang S, Benenson R, Schiele B, et al (2015) Filtered channel features for pedestrian detection. In: CVPR, vol 1, p 4
38. Zinov S, Siemens E (2013) The smart lighting concept. In: Proceeding of the first workshop on problems of autonomous power systems in the Siberian region, Kothen

# Optimization of Control Characteristics Using the Information Model




**Nina Lyubymova, Artem Naumenko, Oleksandr Poliarus, Yevhenii Chepusenko, and Alexandra Lioubimova**

**Abstract** This chapter presents the relevance and necessity of information support modernization for emission management of thermal power plants due to the change in the main technological processes and the transition to new energy sources. This revision of traditional emission control schemes and technologies is necessary to comply with natural resource management standards in Ukraine in the process of its integration into the European system. The study proposes updating environmental standards based on advanced technologies and management experience at thermal power plants in the framework of joint European pilot projects. A mathematical model of information support for the control system has been developed. It is possible to improve the quality control indicators using this information criterion: to increase the performance, reduce the manufacturer's risks, and increase the reliability by 3.5%. In the study, a statistical sample was used for monitoring flue gas pollution of a thermal power plant for 22 days with a 5-min control.

**Keywords** Information support · Emission control · Environment standards · Thermal power plant

N. Lyubymova (✉) · A. Naumenko (✉)
State Biotechnological University, Alchevskykh Street 44, Kharkiv 61000, Ukraine
e-mail: nina.lioubimova@gmail.com

A. Naumenko
e-mail: lajo@ukr.net

O. Poliarus (✉) · Y. Chepusenko (✉)
Department of Metrology and Life Safety, Kharkiv National Automobile and Highway University, Jaroslava Mudrogo Street 25, Kharkiv 61002, Ukraine
e-mail: poliarus.kharkov@ukr.net

Y. Chepusenko
e-mail: yevhenii.chepusenko@gmail.com

A. Lioubimova (✉)
Victoria University of Wellington, 19 Kelburn Parade, Wellington 6012, New Zealand
e-mail: lioubialex@myvuw.ac.nz

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_11

## 1 Introduction and Motivation

Information and management systems (IMS) and environmental control technologies are currently important to carry out multiple tasks for environmental protection and safety under the conditions of a growing anthropogenic pressure from industrial enterprises.

Control of emissions from thermal power plants is one of the key issues of mandatory technological regulations. Unlike environmental control, it must provide concrete results and conclusions. The estimated quality control indicators are based on regulatory documents. The main control indicators should be determined by the established procedures and checked by the subsoil user in accordance with the current environmental legislation. Regulatory documents determine whether such solutions are justified as they establish probable quality control indicators (relevance, reliability, error probability) [1, 2].

## 2 Analysis of Recent Research and Publications

The information support of any control system is closely connected with the object to be controlled in terms of the types and number of informative parameters reflecting basic properties of the object. Environmental pollution control involves a large number of physically heterogeneous parameters reaching hundreds and even thousands of units [3–5].

This leads to certain specific features and difficulties in forming the vector of input signals for information and measurement systems for environmental control (IMS EC), especially for multicomponent pollution processes where the controlled parameters correlate [6].

The challenges associated with environmental pollution by waste from thermal power plants may be exacerbated by the need of transition to new types of energy sources for existing technologies with a shortage of traditional raw materials.

All these factors further strengthen the need of multicomponent emission control modernization for thermal power plants in accordance with the existing requirements of international legislation in order to comply with environmental standards, improving the quality and main characteristics of the process: reliability, speed and reduce risks and the probability of errors. This will optimize basic information and measurement operations and improve their environmental and economic performance.

The increased uncertainty of the controlled parameters of pollution is one of the main factors provoking a decrease in the reliability of environmental control. It is quantitatively manifested in the presence of not excluded components of systematic measurement errors, including the methodological error.

Improving the reliability of control consists, first of all, in improving the information technology of control based on reducing the a priori uncertainty of the probabilistic properties of the control object. This is possible if the planning of measuring experiments for the study of pollution processes will take into account the influence of the time factor on the dynamics of random processes.

This will ensure the adequacy of the probabilistic mathematical model of the pollution process by its physicochemical characteristics of change in time and will allow synthesizing a statistical decision-making model that considers the limited measurement information obtained during the monitoring of the pollution process.

If the requirements for the accuracy and adequacy of the probabilistic model of the pollution process are met and they will ensure the effectiveness of the information control technology, then measurements reducing the uncertainty of the instrumental component associated with increasing the accuracy of measuring instruments will also be effective.

An increase in reliability is equivalent to a decrease in the total probability of a control error, since this probability complements the control reliability to one [5].

The basic methods for increasing the reliability of control, including the analysis of the systematic and random components of control errors probability, are shown in Fig. 1. These methods assume that the systematic component is divided into methodical and instrumental ones [5].

Let's consider in more detail the advantages and disadvantages of individual methods.

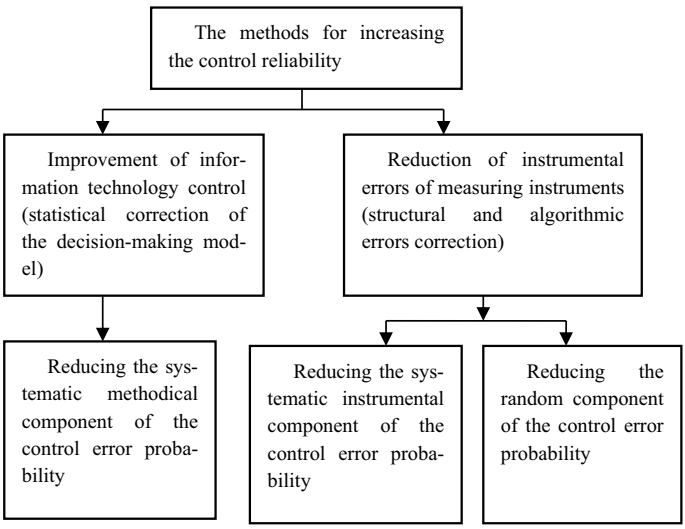



**Fig. 1** Basic methods for increasing the control reliability

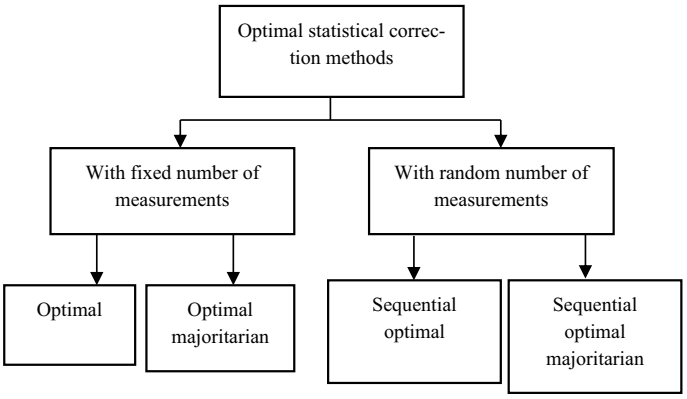



**Fig. 2** Classification of statistical correction methods

Statistical correction methods minimize the average control risk as a linear function of its erroneous decisions under m-component control:

$$\overline{R} = \sum_{j=1}^{m} (q_0 c_j \cdot \alpha_j + q_1 d_j \beta_j), \tag{1}$$

where $q_0$, $q_1$ are the prior probabilities of "no occurrence" and "occurrence" of the norms violations for the maximum permissible emissions (MPE); $c_j$, $d_j$—losses caused by control errors of the first and second type; $\alpha_j$, $\beta_j$—control risks of the first and second type, respectively, for each j-th component of the pollution process.

Correction methods that minimize the average risk (1) are theoretically optimal ones and are based on the use of time redundancy of multiple measurement results of the stationary controlled processes values, provided that their probabilistic models are adequate to a real physical object. Figure 2 describes the types of such methods.

Optimal correction methods are based on the use of likelihood ratio statistics [6]

$$\Lambda = \frac{f_m(x_1, \ldots x_m / s_1)}{f_m(x_1, \ldots x_m / s_0)}, \tag{2}$$

where the numerator and denominator are the conditional m-dimensional likelihood functions of the measured values $x_1, \ldots x_m$ for the m-components for states $s_1$ (object is "not normal") and $s_0$ (object is "normal").

The decisions $\gamma_1$ and $\gamma_0$ taken by the control system are determined by the following relations

$$\begin{cases} \gamma_1 : if \ \ \Lambda \geq 1, \\ \gamma_0 : if \ \ \Lambda < 1 . \end{cases} \tag{3}$$

Statistical correction [7] consists in finding unbiased consistent and effective estimates of constant coefficients for conditional likelihood functions of ratio (2) for an optimal method with a fixed number of measurements.

For the majoritarian optimal correction method $n$ repeated checks are performed, and the selection is carried out for each of them according to the rule (3). If for most of the checks there are $n_1$ solutions $\gamma_1$ and $n_2$ solutions $\gamma_0$ ($n_1 + n_2 = n$), then the final solution is chosen $\gamma_1$ *if* $n_1 > n_2$, and the solution is $\gamma_0$ *if* $n_2 > n_1$. Moreover, $n$ is an odd number.

Sequential optimal methods of statistical correction are based on the use of the Wald criterion [8] of a sequential procedure for analysing $m$ measurement results. In this case, not two, but three solutions are possible:

$$\begin{cases} \gamma_1 : if\, \Lambda \geq (1-\beta)/\alpha, \\ \gamma_0 : if\, \Lambda < \beta/(1-\alpha), \\ \gamma : continue_measuring \\ if\, (1-\beta)/\alpha > \Lambda > (1-\alpha) \end{cases}. \tag{4}$$

The total number of measurements is random, since the statistics $\Lambda$ are random, and the average control duration is the shorter, the greater the set control risks $\alpha$ and $\beta$. In the sequential majoritarian method, two integer comparison thresholds $L$ ($L < n$) and $n - L + 1$ are set for the number $\mu$ of decisions $\gamma_0$. This decision is considered final if $\mu = L$. Otherwise, the decision $\gamma_1$ is considered final one.

The advantage of the considered optimal correction methods for decisions $\gamma_0$ or $\gamma_1$ is the possibility of planning control for the given risks $\alpha$ and $\beta$.

The disadvantage of the methods is that they are mainly applicable for tolerance control [1] in technical diagnostics and control of the technical state of complex industrial products. The disadvantages of statistical correction methods should also include their sensitivity to violations of the model (2) adequacy, if the coefficients of the conditional likelihood functions of the model were incorrectly estimated or their probabilistic properties were incorrectly set.

## 3 The Statement of Considered Problem

One of the main factors leading to a reduced environmental control reliability is an increased uncertainty of controlled parameters in the case of multicomponent environmental pollution (emissions, waste).

Quantitatively, this is due to the presence of not excluded components of systematic measurement errors, including methodological ones [7]. The control reliability can be increased by reducing the a priori uncertainty of the probable properties of the controlled object. For this, in measuring experiments to study the processes of environmental pollution, the influence of time on the dynamics of random processes should be analyzed.

This approach will ensure the adequacy of the probabilistic mathematical model of the pollution process when its physical and chemical characteristics change over time and will allow synthesizing a statistical decision-making model based on limited measurement information obtained in the process of pollution control.

If the requirements for the accuracy and adequacy of the probabilistic contamination model are met, measures for reducing the uncertainty of the instrumental component by increasing the accuracy of measuring instruments will also be effective. An increase in reliability is equivalent to a decrease in the total probability of a control error, since the sum of the probability and reliability of control is equal to 1 [8].

The planning of the control of polluting emissions flow should consider the randomness in time of events in which environmental management norms are violated.

Any technological process violation leads to accidental emissions. The sequence of such events forms a series of events, which are characterized by random times when the specified norms are exceeded [12].

The flow of events is generally a sequence of random points on the time axis with random intervals between them. Such a flow of events is generated by a random process determined by changes of a random variable X (pollution component) in time, any of events in the flow is generated when the process stationarity is disrupted (in terms of mean value, dispersion, spectrum, etc.). Such disruptions cause additional uncertainty in the control of technological pollution which makes probable properties and dynamic features of the processes more complicated.

Planning of the emission flow control must consider not only the sample sizes of the measurement results, but also the order in which they are taken.

The key issue is the choice of a decision-making rule on the basis of a criterion that ensures the given control reliability and guarantees minimization of the risks with considering the level of economic losses in the cases of environmental violations, as well as the potential of using the obtained data to prevent technological accidents (failures).

Purpose and objectives of the study. The aim of the study is to develop an information mathematical model of control for its introduction into practice.

To develop emission control plans for fight with industrial pollution, it is required to solve the following tasks:

- to determine the conditions of measurements, parameters of probabilistic models for the controlled object and the control efficiency indicators that maximize the amount of information obtained due to the control measures;
- to develop a statistical model for control of quantitative exceedance of the norms (MPE, MPD …) by measuring and determining the parameters of the model that minimize uncertainty of the control decisions;
- to explore a probabilistic model of the procedure for emission flow control and develop statistically valid plans for the control measures;

- to investigate probabilistic models of the correlation within the controlled process and develop a statistically valid method for choosing the interval of discrete sampling.

This chapter presents solutions of the individual problems as the scope of chapter is limited.

Figure 3a, b are used to illustrate complexity of the probabilistic structure for multicomponent processes of atmospheric pollution. They describe typical processes of air pollution with emissions from thermal power plants (Fig. 3a, b). Accuracy of registering such exceedances of the norms provided by technological regulations is determined by metrological properties of the measuring instruments and methodological errors of the measurement procedures under specific conditions of measuring experiments.

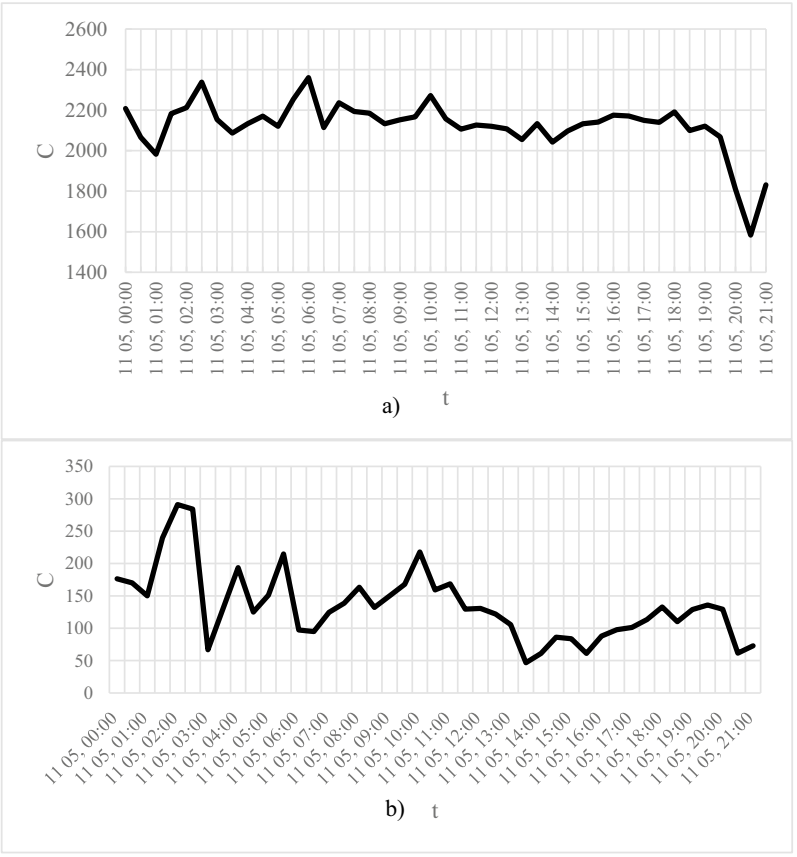



**Fig. 3** Typical realizations of atmospheric pollution by physicochemical components of thermal power plants waste (C—concentration, mg/m$^3$; t—time): **a** $SO_2$; **b** CO

Figure 3a, b shows that random processes of quantitative change in the atmospheric pollution components can be described through complex types of non-stationarity in terms of mathematical expectation, spectrum and the law of probability distribution at the same time. Evaluation of stationarity and spectral features of the pollution processes will reduce a priori uncertainty and will give an opportunity for improving the control planning procedure.

## 4 Use of Information Analysis for Measurement Control of Extreme Emissions Within the Pollution Processes

Measuring control is widely used to establish "the actual value of a parameter relative to its maximum permissible values through the parameter measurement" [1]. Measurement is a basic procedure aimed at receiving primary information about the monitored parameter. The following step is to take a decision on the compliance or non-compliance of the parameter value with the regulatory requirements [13]. The primary measurement information about the actual value of the controlled parameter is further converted into secondary information in the form of logical conclusions (solutions) which gives an opportunity to consider any parametric control system to be an information system.

To determine the amount of information that the system provides, a statistical model is built to establish the criterion for making a decision for the observed process $x(t)$ in the emission presence ($\gamma_1$) or its absence ($\gamma_0$), taking the measured x value as the statistic criterion.

The range of permissible values for the statistics is $\omega_0 \in (0, x_B)$ while the critical range is determined as follows $\overline{\omega} \in [(x_B, \infty]$, where $x_B$ is the MPE norm. The solutions are chosen as follows [14]:

$$\begin{cases} \gamma_0 : x \in \omega_0; \\ \gamma_1 : x \in \overline{\omega}. \end{cases} . \tag{5}$$

In order to describe the probabilistic properties of $X$ in the presence or absence of emissions, the study uses a model of step changes in the mathematical expectation $m_x$ of the $x(t)$ process [15] on the observation interval $\Delta t$

$$\begin{cases} m_x = m, \qquad if\ x(t) \in \omega_0, \\ m_x = m + \Delta\ if\ x(t) \in \overline{\omega}, \Delta = \text{const}, \end{cases} \tag{6}$$

where $\Delta$ as the parameter of displacement determines the parameter of non-stationarity in terms of mathematical expectations.

Two conditions are introduced ($\theta$) for the $x$ process $x(t)$ within the $\Delta t$ interval:

$$\begin{cases} \Theta_0 : x(t) \in \omega_0, \\ \Theta_1 : x(t) \in \overline{\omega}. \end{cases} \tag{7}$$

$X$ can be considered as a sum of a continuous centralized $\overset{\circ}{x}(\Delta t)$ and discrete $Z$ random values

$$x = \overset{\circ}{x}(\Delta t) + Z, \tag{8}$$

where:

$$Z = \begin{cases} m, \quad if \quad \Theta = \Theta_0 \\ m + \Delta, if \quad \Theta = \Theta_1 . \end{cases}$$

Let the $\sigma^2$ dispersion of the $\overset{\circ}{x}$ $(t)$ process is constant for $\Theta_0$ and $\Theta_1$ conditions while *f(x)* is the probability distribution density of the process within the $\Delta t$ interval.

If $\sigma^2 = \text{const}$, for the given conditions (6) and (8) $x(t)$ process can be considered non-stationary one in terms of the mathematical expectation.

Let $T$ is the time observation for the measured process $x(t)$ when the norm $x_B$ is exceeded in the cases of short-term emissions ($T > > \Delta t$).

If the emission flow is stationary, the *H(t)* parameter for the flow will be constant [16]:

$$H(t) = \lambda. \tag{9}$$

If the $\Delta t$ interval is so small that no more than one emission can be placed within this interval, and the emissions are considered as independent random events, the sequence of such events forms the simplest (or stationary Poisson) flow.

For the simplest flow of event, the probability that the exact k number of events will occur within a T time segment is determined according to the following formula [16]:

$$P(k) = \frac{(\lambda T)^k}{k!} \cdot e^{-\lambda T} \tag{10}$$

Equation (10) shows that the a priori probabilities $P_0$ (absence of emissions, k = 0) and $P_1$ (at least one emission) can be determined according to the following formulas [17]:

$$\begin{aligned} P_0 &= e^{-\lambda T}, \\ P_1 &= 1 - e^{-\lambda T}. \end{aligned} \tag{11}$$

## 5 Determination of Expected Information Amount

The uncertainty of X before control is established through differential entropy [18]

$$h_x = \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx. \tag{12}$$

The f(x) density is a combination of two distribution laws [19]

$$f(x) = f(\overset{\circ}{x}) * f(z),$$

where $\overset{\circ}{x}$ is an established $\overset{\circ}{x}(\Delta t)$ value, $z$ is an established $Z$ value at $\overset{\circ}{x} \sim NORM(0, \sigma^2)$, while the $f(z)$ density is set in a distribution range [20]

$$Z : \begin{array}{|c|c|} \hline m & m+\Delta \\ \hline p_0 & p_1 \\ \hline \end{array}$$

Odd central moments of k-order for a discrete random Z variable are as follows [20]:

$$\mu_k = \Delta^k p_0 p_1 (p_0 - p_1). \tag{13}$$

From (8) follows that *f(z)* distribution is symmetrical as well as the normal probability law $f(\overset{\circ}{x})$, if $p_0 = p_1 = 0,5$, as $\mu_k = 0$, for $k = 3, 5, 7, ...k = 3, 5, 7, ...$

The residual differential entropy of X is determined on the basis of the control results in the form of $\gamma_0$ or $\gamma_1$ which corresponds to presence or absence of a stepwise growth $\Delta$ [21] in $z = Z$:

$$h_{x/x_z} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, x_z) \log_2 \frac{f(x, x_z)}{f(x_z)} dx dx_z, \tag{14}$$

where $f(x, x_z)$ is the probability density of co-occurrence for $x$ and $x_z = x(\Delta t) + z$.

The expected information amount about the presence or absence of a random event (emission) according to the results of control is the difference in entropies of $h_x$ and $h_{x/x_z}$ [22]

$$\mathrm{I} = h_x - h_{x/x_z}. \tag{15}$$

If $f(x)$ is a symmetric distribution (normal, even, etc.), according to [23]

$$\mathrm{I} = \log_2 \sqrt{1 + \frac{\sigma_x}{\sigma_{x_z}{}^2}}. \tag{16}$$

the Eq. (16) is valid if the symmetry condition is acceptable for $f(z)$ density, i. e. $p_0 = p_1$. The following step is to establish $\sigma_x^2$ and $\sigma_{x_z}^2$ for an individual case of dispersion.

If the random $z$ and $x(t)$ values are independent, where $t$ is the moment of measuring the $x$ $(t)$ value [24],

$$\sigma_x^2 = (m - m_z)^2 p_0 + (m + \Delta - m_z)^2 p_1 + \sigma^2,$$

where $m_z = m + \Delta p_1$.

After the appropriate transformations, it was established:

$$\sigma_x{}^2 = \Delta^2 p_0 p_1 + \sigma^2. \tag{17}$$

In order to determine the residual dispersion $\sigma_{x_z}{}^2$ let us set $\alpha$ of the first order and $\beta$ of the second order. Denote $\alpha$ and $\beta$ the probabilities of the risks of control of the manufacturer and the customer. Then

$$\begin{aligned} \sigma_{x_z}{}^2 = ({}^m(1-\alpha) + ({}^m\alpha + ({}^m\beta \\ +({}^m \cdot (1-\beta), \end{aligned} \tag{18}$$

where

$$\begin{cases} m_{z|\gamma_0} = m + \Delta\alpha, \\ m_{z|\gamma_1} = m + \Delta(1-\beta). \end{cases} \tag{19}$$

Conditional mathematical expectations $m_{z|\gamma_0}$. and $m_{z|\gamma_1}$ correspond to obtained potential solutions for $\gamma_0$ and $\gamma_1$.

After transformations using the Eqs. (19), the formula (18) will be as follows

$$\sigma_{x_z}{}^2 = \Delta^2[\alpha(1-\alpha) + \beta(1-\beta)]. \tag{20}$$

By substituting the $\sigma_x{}^2$ dispersion values from (17) and (20) in the formula (16) and considering $p_0$ and $p_1$ values in accordance with Eqs. (11), we obtain a more general equation for the expected information to be received during the parametric control of emissions over the T observation time of the pollution process [17].

$$\mathrm{I} = \log_2 \sqrt{1 + \frac{e^{-\lambda T}(1 - e^{-\lambda T}) + \sigma^2/\Delta^2}{\alpha(1-\alpha) + \beta(1-\beta)}}. \tag{21}$$

Equation (21) shows the following [17, 22, 23]:

1. The amount of information is growing with a decrease in the $\alpha$ and $\beta$ control risks. Additionally, analysis of the radical expression shows that maximum of information $I_{max}$ corresponds to $e^{-\lambda T} = 0{,}5$ (i.e. $p_0 = p_1$).
2. The amount of information increases with the parameter of non-stationarity $\Delta$ decrease (change in the mathematical expectation in the $x$ $(t)$ process).
3. Naturally, there is a direct correlation between the amount of information and $\sigma^2$ dispersion of the $x$ $(t)$ process, as the initial entropy grows with an increase of dispersion $h_x$.
4. The amount of information is minimal in case of general control [26], where control risks are as follows $\alpha = \beta = 0{,}5$. Thus, the denominator of the radical expression (21) is maximal, reaching 0,5.
5. It must be noted that Eq. (21) is a local model with $p_0 \cong p_1$ limitations and the condition of stationarity of the emission flow over $T$ observation interval. However, this model can be applied to choose the most efficient observation interval in terms of the maximum expected information as follows:
   $e^{-\lambda T} = 0{,}5$, thus

$$T = -\frac{\ln_2 0,5}{\lambda}. \tag{22}$$

   If $\sigma^2$, $\Delta$, $\alpha$, $\beta$ are unchanged, the amount of information is the highest and is determined on the basis of the following equation [22, 24–27]

$$I = \log_2 \sqrt{1 + \frac{0,25 + \sigma^2/\Delta^2}{\alpha(1-\alpha) + \beta(1-\beta)}}. \tag{23}$$

6. Since the risks of α and β control directly correlate with the measurement errors of the *x (t)* process values [8, 19], a decrease in the number of errors leads to a decline in control risks and, consequently, an increased amount of expected information.

The random processes that are analyzed in the article are, as a rule, non-stationary ones and difficult to analyze. In [28], it was proposed to decompose realizations of non-stationary random processes in an adaptive basis using orthogonal Hilbert-Huang modes, the number of which most often does not exceed 10. As a result, in most cases, Hilbert-Huang modes become stationary ones or close to stationary, which greatly simplifies the analysis.

## 6 Discussion

To check the efficiency of the obtained information model for accidental emissions, a statistical sample of flue gas pollution from TPPs was used.

The result of the practical use of the developed information model was the receipt of additional information, providing an increase in the reliability of the control of a priori non-stationary processes of multicomponent pollution of the air and water environment with waste from energy and energy-intensive enterprises. Also, the purpose of this check was to obtain confirmation of an increase in the speed of control and additional possibilities for predicting abrupt emissions of flue gas pollution.

Such processes are considered as integral elements of the technological chain of energy transformations, the dynamic properties of which are superimposed by random multifactorial effects. The latent regularities of the latter can be used for planning active control of polluting emissions with a given or maximum possible reliability of decision-making with a statistically justified minimization of control risks of both the 1st and 2nd kind [30].

When fossil fuels are burned, waste (flue) gases are generated, which, according to environmental protection rules, cannot be released into the atmosphere without preliminary purification. International and national security authorities in directives and regulations set certain values for the concentration of harmful substances, the values of which must not be exceeded in the flue gas after leaving the chimney [35].

Emissions from coal combustion that can have a significant impact on the environment are primarily carbon dioxide (CO2), sulfur dioxide (SO2), nitrogen oxides (NOx) and dust. Previously, carbon dioxide was not classified as a harmful substance, but today it is purified, since it contributes to the occurrence of the greenhouse effect [30].

The flue gas cleaning measures taken in power plants consist of installations for:

- flue gas sulfur filtration (REA);
- purification from nitrogen oxides (De NOx);
- cleaning flue gas from dust.

There are two alternative options for the purification of waste gases from sulfur and nitrogen oxides, namely, measures that are carried out already during the combustion process (primary measures) and those that act on the resulting flue gas (secondary measures) [36]. In the methods, it is necessary to distinguish the main difference [30]:

- the “with sampling” measurement method is based on the analysis of a sample taken from the process stream and suitably prepared (defined cooling by cooling) outside the process stream atmosphere. In this case, the measurement is carried out under optimal analysis conditions, but with a time delay;
- method of measurements “in place” (In-Situ) means analysis directly in the gas channel, simultaneously with the course of the process and therefore with the possibility of the fastest possible reaction.

However, often the elements of the measuring device are subjected to very harsh process conditions. In addition, as a rule, the measurement takes place in the wet gas of the process, which should be taken into account when comparing the results of the analysis with the results of other methods. Both measurement methods have their own rational areas of application and complement each other.

**Table 1** Device properties and effects of using the ULTROMAT 23

| Properties of the ULTROMAT 23 | Use effect |
|---|---|
| Single-beam method with multilayer detector and integrated automatic calibration function (AUTOCAL) | High selectivity and stability, no span gases required thanks to ambient air calibration |
| Modular design with 1 to 3 infrared channels as an additional option for O2 measurement using a galvanic cell | High cost-effectiveness due to the measurement of up to 4 components in one device and high O2 stability of the cell |
| It is easy to clean the chambers, the O2 cell is very easy to replace | Low maintenance costs |
| SIPROM GA software package for remote management and maintenance. Interface for PROFIBUS PA (optional) | Easy connection to the automation system |

When conducting research on the analysis of the dynamic properties of local changes in the non-stationarity of air pollution processes at a thermal power plant, the data obtained from the statistics of flue gas pollution using a continuous gas analyzer ULTROMAT 23 were used. The properties and effects of using the device are shown in Table 1.

Continuous gas analyzer ULTROMAT 23 is a device operating in sampling systems with simultaneous measurement of CO, NO, SO2, O2 components in one device. An important positive performance of the ULTROMAT 23 is its calibration process using ambient air. A check using expensive test gases is only required once a year.

Table 2 shows the measurement ranges of this device. Its special properties: economical measurement of the main four gas components in one device; high selectivity and accuracy, long-term stability without the use of expensive test gases due to autocalibration in ambient air and others [31].

During the analysis, studies were carried out at the TPP. The experiment lasted 22 days. The total number of multiple measurements for each of the components is 6150 (sampling step—5 min). For multidimensional measurements, a multichannel IMS "Ultramat-23" was used with a maximum measurement error not exceeding 5% (reduced value) [31].

**Table 2** Minimum permissible measurement range for measuring the concentrations of components in emissions from TPPs

| Appliance | Components | Minimum permissible measuring range | |
|---|---|---|---|
| ULTROMAT 23 | CO | 0–150 mg/m$^3$ | 0–250 mg/m$^3$ |
| | $NO_x$ | 0–100 mg/m$^3$ | 0–400 mg/m$^3$ |
| | $SO_2$ | 0–400 mg/m$^3$ | 0–400 mg/m$^3$ |
| | | Instrument for measuring 1–2 components | Instrument for measuring 3 components |

In the process of analyzing the correctness of using the information model, this statistical sample of monitored parameters of flue gas pollution was used.

At the same time, the possibility of predicting abrupt emissions, exceeding the standards, increasing control indicators: speed of response, reduction of manufacturer's risks, approximately by 3.5%, was revealed.

The analysis of the results of the application of the information model with the assessment of the accuracy of the nonstationarity parameter when using the control and warning boundaries for accidental emissions showed improved indicators of the quality of control: increased performance, the ability to predict sudden emissions and their prevention when managing the process of correcting the concentration of pollution in flue gases.

## 7 Conclusion

Experimental studies confirmed the expediency of using information models to improve reliability of control over flue gases from thermal power plants. Use of the proposed information model in reconstruction of power plants will enhance reliability, reduce control-related risks of control and meet the requirements of environmental legislation more efficiently.

## References

1. DSTU 2389-94 (1995) Tekhnichne diagnostuvannya ta kontrol' tekhnichnogo stanu. Termini ta viznachennya. Derzhstandart Ukraine, Kyiv, 23 p.
2. DSTU 2865-94 (1995) Kontrol' nerujnivnij. Termini ta viznachennya. Derzhstandart Ukraine, Kyiv, 52 p.
3. Incecik S, Gertler A, Kassomenos P (2014) Aerosols and air quality. Sci Total Environ 488–489:355. https://doi.org/10.1016/j.scitotenv.2014.04.012
4. Prince O, Ewuzie U, Chibuzo V (2020) Environmental pollution: causes, effects, and the remedies. Microorg Sustain Environ Health 419–429
5. Manisalidis I, Stavropoulou El, Stavropoulos A, Bezirtzoglou E (2020) Environmental and health impacts of air pollution, review. Front Public Health 8:14. https://doi.org/10.3389/fpubh.2020.00014
6. Apostolyuk S, Dzhigirej V (2005) Promislova ekologiya, Navch. Posib. Znanny, Kyiv, 474 p.
7. Kisil I (2000) Metrologiya, tochnist' i nadijnist' zasobiv vimiryuvan, 400 p. Fakel, Iv. Frankovsk. (in Ukraine)
8. Volodars'kij E (2001) Metrologichne zabezpechennya vimiryuvan' i kontrolyu, 219 p. Veles, Vinnicya. (in Ukraine)
9. Kuznecova V, Barzilovicha Y (1990) Nadezhnost' i effektivnost' v tekhnike, Reference book in 10 vols. Mashinostroenie, 320 p. Ekspluataciya i remont, Moskow. (in Russian)
10. Volodarskij E (2008) Statistichna obrobka danih, 308 p. Kyiv, NAU. (in Ukraine)
11. Nazarenko I (2006) Pidvishchennya tochnosti nepryamih elektrichnih metodiv vimiryuvannya vologosti pri vikoristanni metrologichno neviznachenih, vimiryuval'nih signaliv, no 2. Harkiv, Ukraïns'kij metrologichnij zhurnal, pp 54–57. (in Ukraine)
12. Montgomery D (2001) Introduction to statistical quality control. Willey, New York, 796 p.

13. Lyubymova NA, Puzik V, Puzik L (2018) Application of the sequential control for increasing competitiveness of agricultural products. Int J Eng Technol 7(4.3):507–513
14. Malajchuk V, Mozgovij O, Petrenko O (2001) Informacijno-vimiryuval'ni tekhnologiï nerujnivnogo kontrolyu. RVV DNU, Dnipropetrovsk, 240 p. (in Ukraine)
15. Marikuca U (2011) Osoblivosti pobudovi programnogo zabezpechennya sistemi monitoringu navkolishn'ogo seredovishcha, Lviv Politechnic National University Institutional Repository, pp 31–34. http://ena.lp.edu.ua. (in Ukraine)
16. Gihman I, Skorohod J (2012) Teoriya sluchajnyh processov, Kniga potrebovaniyu, Moskow, 664 p. (in Russian)
17. Lyubymova NA (2014) Kontrol' zagryaznyayushchih vybrosov energeticheskih predpriyatij s ispol'zovaniem informacionnogo analiza 3(50):346–349. Visnik HNTU, Herson
18. Sobolev VI (2006) Informacionno-statisticheskaya teoriya izmerenij. Izmerenie sluchajnyh velichin i sluchajnyh vektorov, Training manual. MAI, Moskow, 664 p. (in Russian)
19. Lebedev AV (2011) Teoriya veroyatnostej i matematicheskaya statistika, Training manual. Fadeeva LN (eds) Rid Grupp, Moskow, 496 p. (in Russian)
20. Kudryashov BD (2009) Teoriya informacii. Piter, Saint Petersburg, 322 p. (in Russian)
21. Dushin VK (2010) Teoreticheskie osnovy informacionnyh processov i sistem, Training manual. Dashkov i K, Moskow, 348 p. (in Russian)
22. Sovetov B, Cekhanovskij VV (2005) Informacionnye tekhnologii: uchebnik dlya vuzov, 263 p. Vyssh. shkola, Moskow. (in Russian)
23. Ivanov YUP, Nikitin VG (2011) Informacionno-statisticheskaya teoriya izmerenij. Metody optimal'nogo sinteza informacionno-izmeritel'nyh sistem, kriterii optimizacii i svojstva ocenok, Training manual. GUAP, Saint Petersburg, 160 p. (in Russian)
24. Gvozdeva TV, Ballod BA (2009) Proektirovanie informacionnyh sistem. Feniks, Moskow, p 512. (in Russian)
25. Ivanov YUP, Biryukov BL (2008) Informacionno-statisticheskaya teoriya izmerenij. GUAP, Saint Petersburg, 160 p. (in Russian)
26. Blinkov YUV (2011) Osnovy teorii informacionnyh processov i sistem, Training manual. PGUAS, Penza, 184 p. (in Russian)
27. Box G, Luceno A (1997) Statistical control by monitoring and feedback adjustment. Wiley, New York, 327 p.
28. Poliarus OV, Poliakov Y, Lebedynskyi A (2021) Detection of landmarks by autonomous mobile robots using camera-based sensors in outdoor environments. IEEE Sens J 21(10):11443–11450. https://doi.org/10.1109/JSEN.2020.3010883
29. Balev VN (2015) Metody i sredstva kontrolya pogreshnostej izmeritel'nyh kanalov avtomatizirovannyh sistem kontrolya i upravleniya, dissertation Ph.D. 05.11.15, 216 p. Kharkiv. (in Ukraine)
30. Lyubymova NA (2011) Informacionnoe obespechenie sistem tekhnicheskogo kontrolya processa vozduhoochistki promyshlennyh predpriyatij, Collection of scientific papers. YUZGU, Kursk, pp 188–193. (in Russian)
31. Lyubymova NA (2014) Korrekciya prognoziruyushchih procedur pri kontrole zagryaznyayushchih vybrosov energeticheskih predpriyatij, no. 3/1. Tekhnologichnij audit i rezervi virobnictva, Kharkiv, pp 42–44. (in Ukraine)

# Modern Challenges in Telecommunication Technologies

# Resilience Improvement by Traffic Engineering Fault-Tolerant Routing in Programmable Networks




**Oleksandr Lemeshko, Oleksandra Yeremenko, Maryna Yevdokymenko, Amal Mersni, Valentyn Lemeshko, and Mykhailo Persikov**

**Abstract** The research presents an approach to resilience improvement by Traffic Engineering fault-tolerant routing in programmable networks. The mathematical model allows formalizing software-defined network data plane construction that connects with multiple access networks. Additionally, to increase fault-tolerance, several border routers were utilized. The main aim of the solution is to improve the level of overall network resilience. At the same time, load balancing in the data plane is achieved by applying the Traffic Engineering concept by ensuring the packet flows distribution using primary or backup routes at the access level between several gateways that create one virtual default gateway. The technical task of fault-tolerant routing with load balancing based on a modified model has a linear programming optimization form. The model implements the protection of the default gateway, providing load balancing on the interfaces of the virtual default gateway and within the core network. The numerical research results of Traffic Engineering fault-tolerant routing processes confirmed the proposed model's effectiveness in implementing the default gateway protection scheme and load balancing in the network.

O. Lemeshko (✉) · O. Yeremenko (✉) · M. Yevdokymenko (✉) · A. Mersni (✉) · V. Lemeshko (✉) · M. Persikov (✉)
Kharkiv National University of Radio Electronics, Nauky Avenue 14, Kharkiv 61166, Ukraine
e-mail: oleksandr.lemeshko.ua@ieee.org

O. Yeremenko
e-mail: oleksandra.yeremenko.ua@ieee.org

M. Yevdokymenko
e-mail: maryna.yevdokymenko@ieee.org

A. Mersni
e-mail: amal.mersni@nure.ua

V. Lemeshko
e-mail: valentyn.lemeshko@nure.ua

M. Persikov
e-mail: mykhailo.persikov@nure.ua

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_12

**Keywords** Resilience · Fault-tolerant routing · Traffic engineering · Flow-based model · Border router · SD-WAN

## 1 Introduction

There is no doubt that modern networks are complex technical communication systems. Moreover, networks must maintain their functionality in the face of constant internal and external destructive influences, which in turn cause service failures. Such failures can be provoked by hardware and software failures, overloading, or compromising network equipment [1–3]. Therefore, deploying reliable infocommunication networks is a significant scientific and applied task that requires implementing effective proactive technical solutions when the primary purpose is to prevent incidents that provoke network failures. For their part, reactive mechanisms will provide an appropriate response to possible failures to minimize their negative consequences and maintain the demanded level of Quality of Service (QoS) and network security [3–9].

Among the critical components in the network resilience ensuring, the fault-tolerant routing protocols are distinguished. In particular, the well-known routing protocols such as EIGRP (Enhanced Interior Gateway Routing Protocol) and IS-IS (Intermediate System—to—Intermediate System) formally account for the reliability (namely, errors) of links in metrics when calculating routes [1, 5]. In contrast, when the network link or node (router) failures occur, the Fast ReRouting (FRR) mechanism may be involved, when concurrently with the primary route (or routes in the multipath case adoption), the protocol calculates the backup one [1]. The FRR approach allows redirecting packet flows to backup routes within milli-seconds if the primary paths fail. Practically, recent FRR systems provide essential protection (reservation) methods for network elements and QoS level, namely links, nodes, paths, and bandwidth.

A separate class of fault-tolerant routing implementation in modern networks is the First Hop Redundancy Protocols (FHRP). Such protocols are responsible for network protection from border routers' failures serving as default gateways between access networks and the core. The most popular and widely used protocols associated with the FHRP class include VRRP (Virtual Router Redundancy Protocol), HSRP (Hot Standby Router Protocol), CARP (Common Address Redundancy Protocol), and GLBP (Gateway Load Balancing Protocol) [10–20].

However, the efficiency of network protocols depends on the mathematical models, methodologies, and computing algorithms applied for their implementation. Therefore, improving theoretical solutions in fault-tolerant routing based on the principles of optimum resource use and load balancing seems to be an essential field of study.

## 2 First Hop Redundancy Protocols Overview

Indeed, network redundancy is a mechanism to improve availability and reduce failure at one point, known as a single point of failure (SPOF). Conventionally, SPOF is defined as the defect in the design, setup, or implementation of the system or component that may cause the whole system to fail. SPOF may be eliminated by duplicating and adding links or network components to establish alternative backup systems and preserve network functioning in the event of a failure.

One of the existing approaches is implementing the redundant gateway. The primary problem with gateway redundancy is the router device itself. When the router or link breaks down, access to the external network is lost. As a result, a second redundant router must be deployed to avoid the disconnection of the gateway network. Often the FHRPs inherit the functionality of configuring multiple gateways by enabling hosts to automatically transfer data to the redundant router without requiring manual configuration modifications. Therefore, such protocols overcome the default gateway limitations and provide non-disruptive operation.

In addition to the mentioned above protocols, the following FHRPs have the wide practical application [10–20]:

- Hot Standby Router Protocol (HSRP);
- Virtual Router Redundancy Protocol (VRRP);
- Gateway Load Balancing Protocol (GLBP);
- Common Address Redundancy Protocol (CARP);
- Extreme Standby Router Protocol (ESRP);
- Routed Split multi-link trunking (R-SMLT);
- NetScreen Redundancy Protocol (NSRP);
- Chassis Cluster Redundant Ethernet.

The next subsections are devoted to the basic description of the primary characteristics and improvements of the FHRPs.

### *2.1 HSRP*

The Hot Standby Router Protocol is a Cisco-developed redundancy protocol that configures routers as group members and assigns both virtual IP address (VIP) and MAC address utilized by end devices to communicate with the gateway router [10]. For the HSRP group, the router with the greatest priority is designated as the active router and is responsible for traffic forwarding. The second router with the second-highest priority is elected as the standby router, which will take the responsibility of routing packets in the event of an active router failure or when predefined criteria are fulfilled. All the remaining routers in the HSRP group are in the listening state [11].

HSRPv2 is the protocol's newest version, designed to address the shortcomings of the protocol's default version, HSRPv1. It was enhanced to support IPv6 and use

millisecond `hello` timers. As a result, it offers more excellent stability and a quicker recovery time after outages [5].

Another advantage of HSRPv2 is that it sends `hello` packets using the new IP multicast address 224.0.0.102 instead of the multicast address 224.0.0.2 used by HSRPv1. This new multicast address allows Cisco Group Management Protocol (CGMP) leave processing to be enabled at an equivalent time as HSRP. HSRPv2 increases the group number range to 4096 rather than 256 in HSRPv1. Therefore, the expanded group number range enables the group number on sub-interfaces to correspond to the VLAN number [12].

## *2.2 VRRP*

The Virtual Redundancy Protocol (VRRP) is an open-source fault-tolerant protocol standardized by the Internet Engineering Task Force (IETF) [13]. It is used to provide continuous and reliable network services. In the VRRP group, the active router that controls the VIP gateway is called the master router, while all remaining VRRP routers are known as backup routers. The virtual IP address can be the interface's physical (real) IP address.

The original VRRPv2 was published in RFC 3768 and intended to handle only IPv4. However, the most recent version, VRRPv3, detailed in RFC 5798, supports IPv4 and IPv6 [13]. The benefit of using VRRPv3 is that it allows faster transition to backup devices than standard IPv6 neighbor discovery mechanisms. VRRPv3 allows a backup router to become a master router in a matter of seconds with no overhead traffic and no-host involvement [14]. Another advantage of using VRRPv3 is improving the network redundancy by using several devices as default gateway devices, which excludes the single point of failure.

## *2.3 CARP*

The Common Address Redundancy Protocol (CARP) is a patent-free alternative to Cisco's HSRP and VRRP designed by the OpenBSD developers. It is a gateway redundancy protocol used to enhance the availability of the network and ensure non-interrupted services. It is a secure protocol using the HMAC (Hash-based message authentication code) SHA-1 algorithm (Secure Hash Algorithm 1) and may be deployed on both Pv4 and IPv6 networks [15]. The protocol is used among a group of devices sharing the same IP address on the same network (redundancy group). One host is elected the master within the group, and the rest are known as backups. The master owns the shared IP address and answers any traffic directed at it, such as ARP queries. Also, a host may be a member of several groups [16]. Each node in CARP requires three arguments to function correctly: `advbase`, `advskew`, and `password` [17, 18].

### 2.4 NSRP

NSRP stands for the NetScreen Redundancy Protocol. A router redundancy protocol NSRP owned by Juniper Networks ensures transparent failover and load balancing. For a long time, NSRP was referred to as High Availability (HA). The prior version of NSRP was implemented in earlier generations of NetScreen, for example, the NS-100, which does not support ScreenOS 4.0 or later. This version is restricted in functionality and needs additional commands to set up HA. The release of version 2 of the NSRP is implemented in ScreenOS 3.1. HA setup is often known as NSRP configuration or "NSRP operation" [19].

### 2.5 GLBP

The Gateway Load-Balancing Protocol (GLBP) is another Cisco Proprietary redundancy protocol. It enables load sharing of packets across a set of redundant routers. GLBP guarantees load balancing capability for multiple gateway routers with the same IP address but different MAC addresses. In GLBP, all routers operate as active routers to prevent the whole system from collapsing. It improves network performance by load balancing [10–12].

In HSRP and VRRP groups, just one router handles all active traffic, leaving the other routers inactive, resulting in a wastage of available resources. This drawback, however, may be mitigated via further administrative configurations that enable load sharing across hosts belonging to different groups. Therefore, GLBP is founded with the primary objective of resolving all of these concerns. The protocol allows the creation of up to 1024 GLBP groups in every router's physical interface. Roles in the GLBP group are also defined by devices' priorities, with the rule of the highest priority gaining precedence. There are two leading roles of gateways, one Active Virtual Gateway (AVG) and up to four Active Virtual Forwarders (AVF) per group. The AVG replies to ARP requests for the default gateway's MAC address. At the same time, the AVR is in charge of traffic routing [10, 20].

### 2.6 Analysis of Fault-Tolerant SDN Solutions

Many researchers are engaged in studying the problem of improving and developing new solutions of fault-tolerance capability in Software-Defined Networks (SDNs) [21–28]. Hence, several studies have been conducted and published recently. In [23], an approach to deal with the links failures in the data plane named Controlled based Robust Network (CORONET) was suggested. It is based on the NOX platform, which enables interaction with open flow switches. While [24] has proposed BOND, a flexible failure recovery mechanism in SDNs. The idea is to assign backup rules

to forwarding switches rather than backup routes while accounting for flow needs. Also, a global hash table to expedite failure recovery is used. Extensive experiments applying real-world topologies were built to illustrate BOND's efficiency.

In [25], FTLink is proposed as an effective link fault tolerance method in SDN. It creates backup links if the primary connection fails. The FTLink generates a matching table for the backup rule maintained through the controller. The Byzantine Fault Tolerant (BFT) distributed SDN controller was developed in work [26]. It is a suggested prototype capable of tolerating Byzantine errors in data and control SDN planes. The BFT mechanism combines two open-source Byzantine Fault vulnerable SDN controllers with a Byzantine fault-tolerant state machine replication software suite.

However, the previous proposed studies have significant limitations. Indeed, some solutions are only concerned with the data plane link failures, such as CORONET, BOND, and FTLink. Whereas AFRO, FT-SDN, LBFT, MORPH deal with the SDN control plane's failures [21, 22, 27, 28]. As a result, additional defect categories become evident as the SDN architecture develops. Thorough evaluations and continuous improvement are needed to stay on the cutting edge of the SDN fault tolerance aspect.

## 3 Traffic Engineering Fault-Tolerant Routing Flow-Based Model

In the Traffic Engineering fault-tolerant routing flow-based model, we assume that the graph $G = (M, L)$ describes the network structure where the following parameters defined in Table 1. Then $K$ is the set of flows emanating from the access networks and reaching the core network. Let us consider the $k$th flow ($k \in K$) associated with $V_s^k$ and $V_d^k$ (Table 1). Both parameters represent the source and destination access networks for the $k$th flow. The parameter $\lambda^k$ is the average packet rate, also referred to as the $k$th flow intensity at the entrance to the network measured in packets per second (pps).

The set $R$ consists of two disjoint subsets. The first subset of vertices $R^+$ simulates routers residing at the boundary of the network core and may be linked to the access networks. The second subset $R^-$ represents transit routers. Furthermore, a subset $R_j^+$ of the $R^+$ simulating edge routers interfaces comprises a virtual router for the $j$th access network specified by $V_j$ vertex. In turn, $|R_j^+| = m_j^+$ is the size of a subset of edge routers interfaces required to create a virtual router for the $j$th access network. It is worth noting that interfaces on a single edge router may simultaneously belong to several virtual routers.

In addition to that, the routers set $R_i^{core}$ presents the core network part of neighbors with the router $R_i \in R$. Also, the set $R_i^{access}$ defines access networks adjacent to the border (edge) router $R_i \in R^+$.

**Table 1** Notation summary

| Symbol | Meaning |
|---|---|
| $G = (M, L)$ | Network graph |
| $M = R \cup V$ | Set of vertices ($R \cap V = \varnothing$) |
| $R = \left\{R_i, i = \overline{1, m}\right\}$ | Core network routers |
| $V = \left\{V_j, j = \overline{1, v}\right\}$ | Access networks |
| $R^+ \subset R$ | Core network edge routers |
| $R^- \subset R$ | Transit routers |
| $R_j^+ \subset R^+$ | Edge routers that form a virtual router for the $j$th access network $V_j$ |
| $L = E \cup W$ | Set of edges ($E \cap W = \varnothing$) |
| $E = \left\{E_{i,j}, i, j = \overline{1, m}, i \neq j\right\}$ | Core network links |
| $W = \left\{W_{i,j}, i = \overline{1, v}, j = \overline{1, m^+}\right\}$ | Access links for access and core networks connection |
| $\varphi_{i,j}$ | Link $E_{i,j} \in E$ capacity |
| $R_i^{core}$ | Core network routers adjacent to the border router $R_i \in R$ |
| $R_i^{access}$ | Access networks adjacent to the border router $R_i \in R^+$ |
| $K$ | Set of flows arriving at the edge routers from the access networks |
| $K_i^+$ | Set of flows arriving at the core from the access networks connected to the router $R_i \in R^+$ |
| $K_i^-$ | Set of flows outgoing from the core to the access networks connected to the router $R_i \in R^+$ |
| $V_s^k$ | Access network that is the source for $k$th flow |
| $V_d^k$ | Access network that is the destination for $k$th flow |
| $\lambda^k$ | $k$th flow average packet rate |
| $x_{i,j}^k$ | $k$th flow fraction in the link $E_{i,j}$ of the primary link |
| $y_{i,j}^k$ | $k$th flow fraction in a primary access link $W_{i,j}$ |
| $z_{j,i}^k$ | $k$th flow fraction in a primary access link $W_{j,i}$ |
| $\overline{x}_{i,j}^k$ | $k$th flow fraction in the link $E_{i,j}$ of the backup path |
| $\overline{y}_{i,j}^k$ | $k$th flow fraction in a backup access link $W_{i,j}$ |
| $\overline{z}_{j,i}^k$ | $k$th flow fraction in a backup access link $W_{j,i}$ |
| $\alpha$ | Upper bound of the network links utilization |

The following basic conditions must be fulfilled to implement the multipath routing [29–31]:

$$0 \leq x_{i,j}^k \leq 1. \tag{1}$$

Applying a multipath routing strategy in compliance with Traffic Engineering technology guarantees load balancing and improves the overall Quality of Service in the network [2, 3].

The model contains the following conditions under the possibility of load balancing over virtual router interfaces by analogy with HSRP and GLBP [30, 31]:

$$0 \le y_{i,j}^k \le 1 \text{ and } 0 \le z_{j,i}^k \le 1. \tag{2}$$

Moreover, the subsequent conditions take place to prevent packet losses at the core network edge:

$$\sum_{R_j \in R_p^+} y_{p,j}^k = 1, V_p = V_s^k; \tag{3}$$

$$\sum_{R_j \in R_h^+} z_{j,h}^k = 1, \quad V_h = V_d^k. \tag{4}$$

Then the flow conservation conditions are imposed on routing variables [30], the fulfillment of which allows interaction in calculating three types of control variables and coordinate processes of loading balancing at the core and edge network levels:

$$\begin{cases} \sum\limits_{j:E_{i,j} \in E} x_{i,j}^k - \sum\limits_{j:E_{j,i} \in E} x_{j,i}^k = 0; k \in K, R_i \in R^-; \\ \sum\limits_{j:E_{i,j} \in E} x_{i,j}^k - \sum\limits_{j:E_{j,i} \in E} x_{j,i}^k = y_{p,i}^k; k \in K, R_i \in R^+, V_p = V_s^k; \\ \sum\limits_{j:E_{i,j} \in E} x_{i,j}^k - \sum\limits_{j:E_{j,i} \in E} x_{j,i}^k = -z_{i,h}^k; k \in K, R_i \in R^+, V_h = V_d^k. \end{cases} \tag{5}$$

Next, it is necessary to introduce additional control variables that determine backup paths [30, 31] to improve routing fault-tolerance when access networks are connected with the core via specific virtual router interface(s): $\overline{x}_{i,j}^k$, $\overline{y}_{i,j}^k$, and $\overline{z}_{j,i}^k$ (Table 1).

The conditions (1)–(5) are also imposed on backup routing variables, and flow conservation conditions take the form:

$$\begin{cases} \sum\limits_{j:E_{i,j} \in E} \overline{x}_{i,j}^k - \sum\limits_{j:E_{j,i} \in E} \overline{x}_{j,i}^k = 0; k \in K, R_i \in R^-; \\ \sum\limits_{j:E_{i,j} \in E} \overline{x}_{i,j}^k - \sum\limits_{j:E_{j,i} \in E} \overline{x}_{j,i}^k = \overline{y}_{p,i}^k; k \in K, R_i \in R^+, V_p = V_s^k; \\ \sum\limits_{j:E_{i,j} \in E} \overline{x}_{i,j}^k - \sum\limits_{j:E_{j,i} \in E} \overline{x}_{j,i}^k = -\overline{z}_{i,h}^k; k \in K, R_i \in R^+, V_h = V_d^k. \end{cases} \tag{6}$$

The protection conditions for the border router $R_i$ are as follows [31]:

- for flows that arrive at the core network via $R_i$

$$0 \le \overline{x}_{i,j}^k \le \delta_{i,j}^+ \text{ under } R_j \in R_i^{core}, \tag{7}$$

$$\delta_{i,j}^{+} = \begin{cases} 0, & \text{when protecting the router } R_i; \\ 1, & \text{otherwise.} \end{cases} \tag{8}$$

- for flows that outgo from the core network via $R_i$

$$0 \le \overline{x}_{j,i}^{k} \le \delta_{j,i}^{-} \text{ under } R_j \in R_i^{core}, \tag{9}$$

$$\delta_{j,i}^{-} = \begin{cases} 0, & \text{when protecting the router } R_i; \\ 1, & \text{otherwise.} \end{cases} \tag{10}$$

Conditions (7)–(10) fulfillment guarantees the border router $R_i \in R$ protection. Under (7)–(10), the backup route cannot utilize all outgoing links from the protecting node.

As shown in [3], to guarantee load distribution throughout the routing process in the network, under the concepts of Traffic Engineering in the proposed model, the criteria of overload prevention are introduced:

- for primary (main) routes

$$\sum_{k \in K} \lambda^k x_{i,j}^k \le \alpha \varphi_{i,j},\ E_{i,j} \in E; \tag{11}$$

- for backup (alternative) routes

$$\sum_{k \in K} \lambda^k \overline{x}_{i,j}^k \le \overline{\alpha} \varphi_{i,j},\ E_{i,j} \in E. \tag{12}$$

The control variables $\alpha$ and $\overline{\alpha}$ define the upper bound of the network link utilization for primary and backup paths accordingly. These variables comply with the constraints [3]:

$$0 \le \alpha \le 1, 0 \le \overline{\alpha} \le 1. \tag{13}$$

To solve the task of the Traffic Engineering fault-tolerant routing, the minimum of the boundary values $\alpha$ and $\overline{\alpha}$ serves as the optimality criterion [30]:

$$\min_{x,y,z,\alpha,\overline{x},\overline{y},\overline{z},\overline{\alpha}} (\alpha + \overline{\alpha}). \tag{14}$$

Within the presented model (1)–(14), the technical task of Traffic Engineering fault-tolerant routing has been set as the linear programming optimization problem with optimality criterion (14), as well as constraints and conditions for control variables (1)–(13).

## 4 Numerical Research and Evaluation of the Traffic Engineering Fault-Tolerant Routing

In this section, we study the processes of Traffic Engineering fault-tolerant routing. Figure 1 displays the network structure for numerical research. Gaps of communication links indicate their bandwidth in packets per second unit.

Suppose two packet flows with intensities $\lambda^1$ and $\lambda^2$, the values of which varied from 10 to 700 pps with a step of 10 pps, are transmitted from the access network $V_1$ to $V_2$ (Fig. 1). During the study, the bounds $\alpha$ and $\overline{\alpha}$ were estimated when implementing the default gateway protection scheme in a border router $R_4$ failure.

Then Fig. 2 shows the dependence $\alpha$ on the intensity values $\lambda^1$ and $\lambda^2$ when using the primary routes for a balancing option similar to the GLBP protocol:

$$y_{i,j}^k \in \{0; 1\} \text{ and } z_{i,j}^k \in \{0; 1\} \tag{15}$$

The fulfillment of conditions (15) ensures load balancing not by packets but by hosts, i.e., flow packets originating from a specific host can be routed to a single border router. In general, hosts can use different border routers whose interfaces form the default virtual gateway.

Figure 3 shows the dependence on the intensity values $\lambda^1$ and $\lambda^2$ when using the primary routes for the balancing option under the proposed model (1)–(14). In both cases, the highest values $\alpha$ correspond to the flows' intensities with approximately the same high range values.

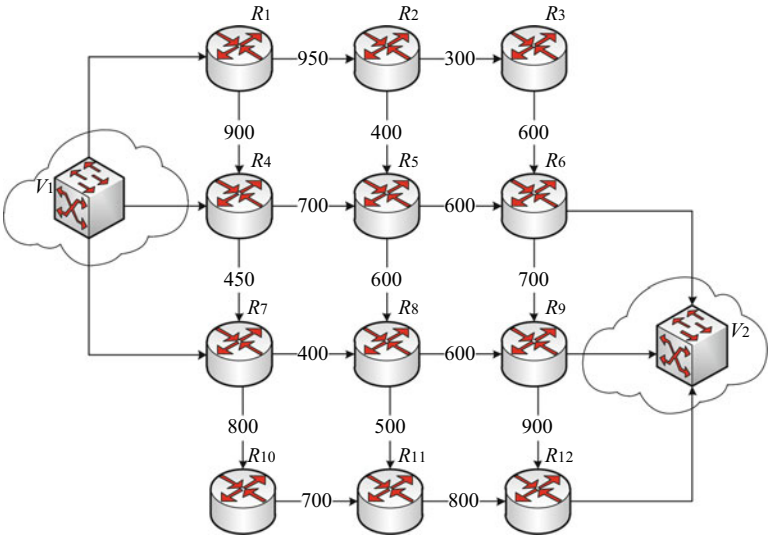



**Fig. 1** Network structure under the numerical study

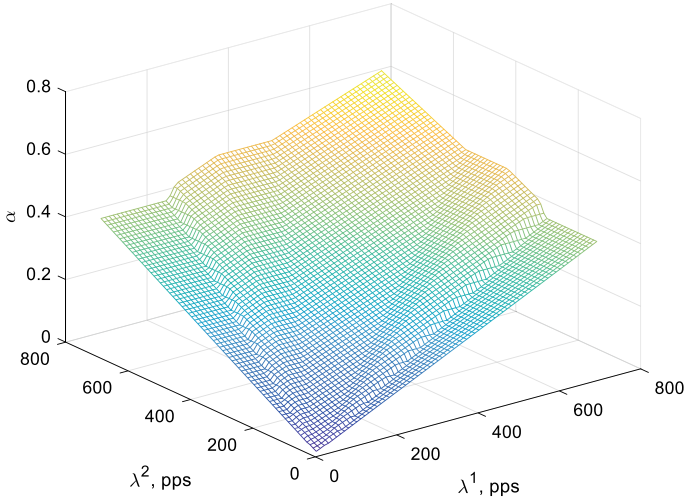



**Fig. 2** Dependence α on the intensities $\lambda^1$ and $\lambda^2$ when using the primary routes for a balancing option similar to the GLBP protocol (15)

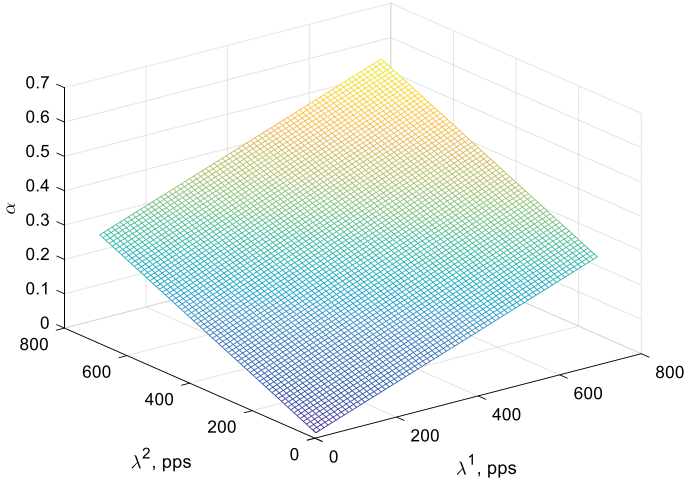



**Fig. 3** Dependence α on the intensities $\lambda^1$ and $\lambda^2$ when using the primary routes for a balancing option according to the proposed model (1)–(14)

Therefore, Fig. 4 shows the gain (in percents) in terms of the boundary value α obtained from using the proposed model (1)–(14) in comparison with solutions similar to the GLBP protocol (15) in the case of using the primary routes.

Consequently, applying the proposed Traffic Engineering fault-tolerant routing mathematical model with primary routes reduces the upper bound of the links utilization α from 10 to 25% in the region of commensurate flows intensities arriving at the network, and up to 40% otherwise (Fig. 4).

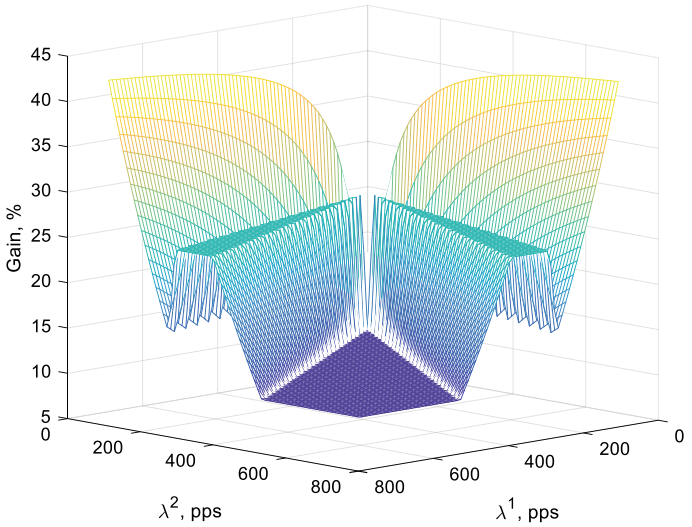



**Fig. 4** The gain (in percents) in terms of α obtained from using the proposed model (1)–(14) in comparison with GLBP solutions (15) using the primary routes

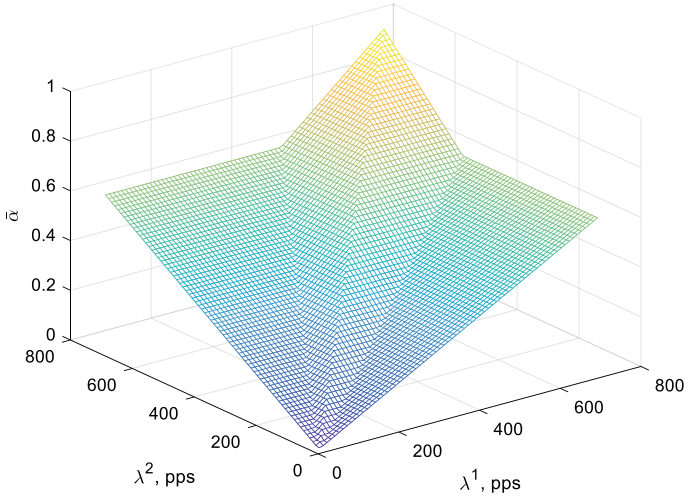



**Fig. 5** Dependence $\overline{\alpha}$ on the intensities $\lambda^1$ and $\lambda^2$ when using the backup routes for a balancing option similar to the GLBP protocol (15)

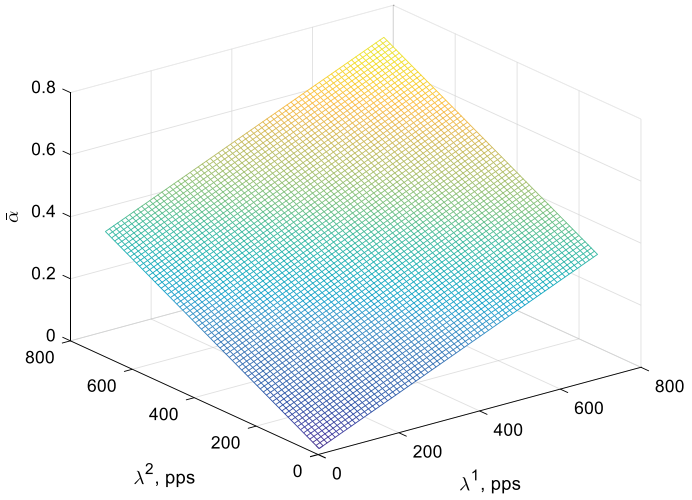



**Fig. 6** Dependence $\bar{\alpha}$ on the intensities $\lambda^1$ and $\lambda^2$ when using the backup routes for a balancing option according to the proposed model (1)–(14)

In the event of a border router $R_4$ failure and redundant routes, network load balancing will change somewhat. The dependences of the bound $\bar{\alpha}$ on the intensities $\lambda^1$ and $\lambda^2$ when using backup routes under load balancing utilizing the GLBP protocol (Fig. 5) and the proposed model (1)–(14) (Fig. 6) were obtained.

Also, Fig. 5 demonstrates that the highest values $\bar{\alpha}$ correspond to the flows intensities that have approximately the same values.

Then Fig. 7 shows the gain (in percents) in terms of the boundary value $\bar{\alpha}$ obtained from using the proposed model (1)–(14) in comparison with solutions similar to the GLBP protocol (15) in the case of using backup routes.

Therefore, applying the proposed Traffic Engineering fault-tolerant routing mathematical model with backup routes allows reducing the upper bound of the network links utilization $\bar{\alpha}$ up to 30% in the region of commensurate flows intensities arriving at the network, and up to 60% otherwise (Fig. 7).

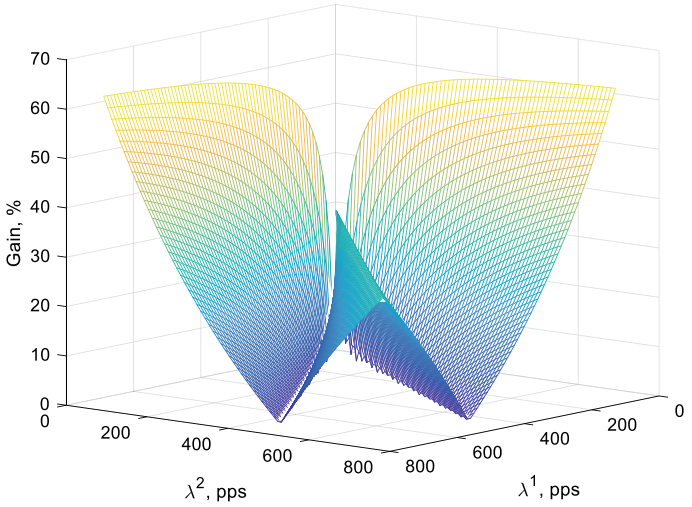



**Fig. 7** The gain (in percents) in terms of $\overline{\alpha}$ obtained from using the proposed model (1)–(14) in comparison with GLBP solutions (15) using the primary routes

## 5 Resilience Aware Traffic Engineering Flow-Based Model

Consider within the model (1)–(10) that load balancing of traffic coming from the access network $V_p$ is implemented, for example, using the GLBP ($y^k_{p,j} \in \{0; 1\}$) by fulfilling the conditions [31]:

$$\sum_{k \in K_j^+} \lambda^k y^k_{p,j} = m^+_{p,j} \sum_{k \in K_j^+} \lambda^k, \tag{16}$$

where $m^+_{p,j}$ are balancing metrics that determine the share of aggregate traffic coming from the access network $V_p$ to the border router $R_j \in R_p^+$. That is, equality must be observed when determining balancing metrics

$$\sum_{R_j \in R_p^+} m^+_{p,j} = 1. \tag{17}$$

In the case of uniform load balancing (Round-robin load-balancing), we have:

$$m^+_{p,j} = \frac{1}{\left|R_p^+\right|}. \tag{18}$$

Therefore, the balancing metrics for the individual border routers that create the default virtual gateway for the access network $V_p$ will be the same.

If the network supports balancing not only at the level of individual flows when $y_{p,j}^k \in \{0; 1\}$ but also at the level of packets of each flow separately under condition $0 \le y_{i,j}^k \le 1$, then one of the available solutions can be considered

$$y_{i,j}^k = \frac{1}{\left|R_i^+\right|}. \tag{19}$$

In addition, the GLBP protocol supports weighted load-balancing, which in a model (16) corresponds to the administrative establishing of balancing metrics. This study also proposes implementing weighted load balancing by adapting the balancing metrics according to the reliability of the border routers, whose interfaces are used for access networks as default gateways.

Let an availability factor $A_i$ characterize each border router $R_i \in R^+$ according to its level of reliability. It is defined as the ratio of the time when the router was operational to the total time of its operation, i.e., takes a value from zero to one. Therefore, to implement resilience aware load-balancing in the system (16), the balancing metrics are proposed to be determined by the following formula:

$$m_{p,j}^+ = \frac{A_j}{\sum\limits_{R_i \in R_p^+} A_i}, R_j \in R_p^+. \tag{20}$$

Thus, in the case of load balancing between the virtual router interfaces, more packets will be sent to a more reliable network device. In the event of a failure of one border router $R_s$, for example, its availability factor becomes zero, and the variables $\overline{y}_{p,j}^k$ must meet the following conditions:

$$\sum_{k \in K_j^+} \lambda^k \overline{y}_{p,j}^k = \overline{m}_{p,j}^+ \sum_{k \in K_j^+} \lambda^k, \tag{21}$$

when the balancing metrics for the backup solution are calculated as

$$\overline{m}_{p,j}^+ = \frac{A_j}{\sum\limits_{R_i \in \overline{R}_p^+} A_i}, R_j \in \overline{R}_p^+, \tag{22}$$

where $\overline{R}_p^+ = R_p^+ \backslash R_s$.

The solution represented by expressions (16)–(22) refers to the border routers through which traffic arrives at the core network. For the case of load balancing on border routers $R_j \in R_p^+$, through which traffic outgoes from the core network, the following conditions apply [31]:

$$\sum_{k \in K_j^-} \lambda^k z_{j,p}^k = m_{j,p}^- \sum_{k \in K_j^-} \lambda^k, \tag{23}$$

$$m_{j,p}^{-} = \frac{A_j}{\sum\limits_{R_i \in R_p^+} A_i}, R_j \in R_p^+, \tag{24}$$

where $m_{j,p}^{-}$ are balancing metrics that determine the proportion of aggregate traffic arriving at the access network $V_p$ from the border router $R_j \in R_p^+$.

Similarly, if the border router $R_s$ fails, its availability factor will be zero, and conditions (23) and (24) for the backup routing solution are supplemented by the following conditions [31]:

$$\sum_{k \in K_j^-} \lambda^k \overline{z}_{j,p}^k = \overline{m}_{j,p}^{-} \sum_{k \in K_j^-} \lambda^k, \tag{25}$$

$$\overline{m}_{j,p}^{-} = \frac{A_j}{\sum\limits_{R_i \in \overline{R}_p^+} A_i}, R_j \in \overline{R}_p^+, \tag{26}$$

where $\overline{m}_{j,p}^{-}$ are load balancing metrics at the core network out for the backup routing solution.

Load balancing between network routers is proposed to be implemented, as shown in [3], by ensuring compliance with the condition of overload preventing:

$$\sum_{k \in K} \lambda^k \max\left(x_{i,j}^k, \overline{x}_{i,j}^k\right) \le \alpha \varphi_{i,j}, E_{i,j} \in E. \tag{27}$$

where $\alpha$ is the control variable that numerically determines the upper bound of the network links utilization (for primary and backup paths) and obeys the constraint (13).

Finally, as the optimality criterion in the problem solving of the Resilience Aware Traffic Engineering (RATE) fault-tolerant routing technological task, the minimum of the boundary value $\alpha$ is selected:

$$\min_{x,y,z,\alpha,\overline{x},\overline{y},\overline{z}} \alpha. \tag{28}$$

Conditions (1)–(10), (13), (16), (17), (21), (23), (25), and (27) act as constraints on control variables when formulating an optimization problem with the optimality criterion (28).

## 6 Numerical Research of the RATE Model

Figure 1 shows the network structure used during the numerical research. As before, the network consists of twelve routers, and communication link breaks indicate their

**Table 2** Availability factors of border routers

| Routers | $R_1$ | $R_4$ | $R_7$ |
|---|---|---|---|
| Variant 1 | | | |
| Availability factor | 0.99 | 0.9 | 0.8 |
| Balancing metric | 0.3680 | 0.3346 | 0.2974 |
| Variant 2 | | | |
| Availability factor | 0.99 | 0.9 | 0.4 |
| Balancing metric | 0.4323 | 0.3930 | 0.1747 |

bandwidth. One packet flow was transmitted between two access networks. The packet flow intensity from the first to the second access network was 2100 pps. The first access network used the interfaces of routers $R_1$, $R_4$, and $R_7$ as a virtual gateway. For the second access network, the default virtual gateway was created by the border routers $R_6$, $R_9$, and $R_{12}$ interfaces. The solutions obtained using the Traffic Engineering FHRP model (TE-model) considered in [30], based on the use of expressions (1)–(10), (13), (27), (28), and the proposed in this work RATE-model (1)–(10), (13), (16)–(28), i.e., with the introduction of additional conditions (16)–(26). Solutions in the event of the $R_7$ router failure were analyzed and compared. The efficiency of routing solutions was evaluated by the value of the upper bound of the network links utilization $\alpha$ for different variants of the border routers' availability (Table 2).

As a result of research, it was found that the introduction of balancing metrics can significantly change the order of balancing traffic arriving from access networks to border routers. The more reliable the border router, the more load will be directed to its interfaces from access networks. It should be noted that with a slight difference in the reliability of border routers (Table 2, Variant 1), consideration within the RATE-model of $A_i$ parameters did not lead to $\alpha$ deterioration, i.e., the QoS level in the network (Table 3). Within the numerical example, the upper bound of the network links utilization remained at the level of 0.9130. In Table 3, the gray color highlights those links that meet the upper bound values of link utilization which is $\alpha = 0.9130$. Therefore, the network resource in the implementation of the Traffic Engineering principles was enough to ensure effective load balancing. This result shows a significant advantage of the proposed solution presented in the RATE model.

In some cases, the increase in the upper bound $\alpha$ occurred only when the difference in the values of the availability factor of the routers by about two or more times (Table 3, Variant 2). Within the example, the link utilization upper bound increased to 0.9369 (Table 3). It means that the optimal from the point of view of the reliability of border routers load balancing made some adjustments to the work of Traffic Engineering in the network in general. However, this situation is not typical for the case where failure is caused by a malfunction of network equipment only. As a rule, the reliability of network core routers is measured by availability factors with four or five nines after the comma. Therefore, the scenario with a sharp decrease in availability factors and a worsening of the upper bound (13) can be caused only by external factors, such as intrusions and network attacks to exploit vulnerabilities on routers, etc. [5].

**Table 3** Comparison of routing solutions obtained using TE and RATE models (without failure of $R_7$ router)

| Link | TE-model | RATE-model | |
|---|---|---|---|
| | | Variant 1 | Variant 2 |
| | $\alpha = 0.9130$ | | $\alpha = 0.9369$ |
| $W_{1,2}$ | 639.1304 | 772.8625 | 907.8603 |
| $W_{1,4}$ | 639.1304 | 702.6022 | 825.3275 |
| $W_{1,7}$ | 821.7391 | 624.5353 | 366.8122 |
| $E_{1,2}$ | 639.1304 | 639.1304 | 655.8008 |
| $E_{2,3}$ | 273.9130 | 273.9130 | 281.0575 |
| $E_{1,4}$ | 0 | 133.7320 | 252.0595 |
| $E_{2,5}$ | 365.2174 | 365.2174 | 374.7433 |
| $E_{3,6}$ | 273.9130 | 273.9130 | 281.0575 |
| $E_{4,5}$ | 639.1304 | 456.5217 | 655.8008 |
| $E_{5,6}$ | 547.8261 | 547.8261 | 562.1150 |
| $E_{4,7}$ | 0 | 379.8125 | 421.5862 |
| $E_{5,8}$ | 456.5217 | 273.9130 | 468.4291 |
| $E_{6,9}$ | 0 | 0 | 0 |
| $E_{7,8}$ | 365.2174 | 365.2174 | 132.5977 |
| $E_{8,9}$ | 547.8261 | 547.8261 | 562.1150 |
| $E_{7,10}$ | 456.5217 | 639.1304 | 655.8008 |
| $E_{8,11}$ | 273.9130 | 91.3043 | 38.9118 |
| $E_{9,12}$ | 0 | 0 | 0 |
| $E_{10,11}$ | 456.5217 | 639.1304 | 655.8008 |
| $E_{11,12}$ | 730.4348 | 730.4348 | 694.7126 |
| $W_{6,2}$ | 821.7391 | 821.7391 | 843.1724 |
| $W_{9,2}$ | 547.8261 | 547.8261 | 562.1150 |
| $W_{12,2}$ | 730.4348 | 730.4348 | 694.7126 |

In the case of the $R_7$ border router failure, the use of the entire network segment was blocked, which led to a decrease in network bandwidth in general. However, even in conditions of reduced available bandwidth, as in the previous case (Table 3), the introduction of the RATE model did not increase the upper bound (13) compared to the TE model use.

## 7 Conclusion

In this work, the investigation of an enhanced mathematical model for Traffic Engineering fault-tolerant routing was carried out. The suggested model of fault-tolerant routing is flow-based and represented by mathematical expressions (1)–(14). The model covers various options for load balancing in the network core (1), (5) and at the access level (2), (5). In numerical research example, conditions (15) allow us to describe the features of load balancing between border routers that form a virtual default gateway based on the principles of the GLBP protocol.

According to the Traffic Engineering principles, the choice of the optimality criterion (14) organizes load balancing in the network core and access level. More specifically, it minimizes the upper bound of the network links utilization. Additionally, the model covers the implementation cases of default gateway protection schemes, as well as routers and links of the network core (6)–(8) protection.

The conducted numerical research establishes the model effectiveness for Traffic Engineering fault-tolerant routing. Its application for a specific network topology (Fig. 1) allows improvement of the routing solutions fault tolerance and reduces the upper bound of the network links utilization:

- from 10–25% to 40% when using the primary routes (Fig. 4);
- from 30 to 60% when using backup routes (Fig. 7).

Reducing the upper bound of the network links utilization contributes to improving the network performance, average end-to-end packet delay, jitter, and packet loss probability, which are the leading QoS indicators.

The next improvement was developing the resilience aware Traffic Engineering (RATE) fault-tolerant routing task. The RATE-model (1)–(10), (13), (16), (21), (23), (25), and (27) of fault-tolerant routing with default gateway protection ensures load balancing between border routers. Within the proposed model, the problem of fault-tolerant routing is formulated in an optimization form with the optimality criterion (28) and constraints (1)–(10), (13), (16), (21), (23), (25), and (27). The novelty of the proposed model of fault-tolerant routing is the introduction of conditions (16)–(26) to ensure the balancing of the load arriving from the access networks between the border routers, considering their availability. That is, the less reliable the border router is, and the higher its failure probability, the less load from the access networks will be sent to this router. This mechanism will minimize the number of packets lost in the event of the actual failure of the least reliable network border routers.

The numerical research confirmed the efficiency of the network solutions received utilizing the offered fault-tolerant routing model. It is established that with a slight difference in the reliability of border routers, consideration within the RATE-model of $A_i$ parameters did not lead to $\alpha$ deterioration and the QoS level degradation in the network in general (Table 3). However, only in some cases did an increase (up to 3%) in the upper bound occur with a difference in the routers' availability factors by about two or more times, which is not typical for modern networks.

The directions for future research are related to the extension of the parameters to be protected within the fault-tolerant routing processes to Quality of Service and network security indicators.

## References


1. Chiesa M, Kamisiński A, Rak J, Rétvári G, Schmid S (2021) A survey of fast-recovery mechanisms in packet-switched networks. IEEE Commun Surv Tutor 23(2):1253–1301. https://doi.org/10.1109/COMST.2021.3063980
2. Tomovic S, Radusinovic I (2018) A new traffic engineering approach for QoS provisioning and failure recovery in SDN-based ISP networks. In: 2018 23rd international scientific-professional conference on information technology (IT) proceedings. IEEE, pp 1–4. https://doi.org/10.1109/SPIT.2018.8350854
3. Lemeshko O, Yeremenko O, Hailan AM (2017) Two-level method of fast ReRouting in software-defined networks. In: 2017 4th international scientific-practical conference problems of infocommunications. Science and technology (PIC S&T) proceedings. IEEE, pp 376–379. https://doi.org/10.1109/INFOCOMMST.2017.8246420
4. Lemeshko O, Yevdokymenko M, Yeremenko O, Hailan AM, Segeč P, Papán J (2019) Design of the fast reroute QoS protection scheme for bandwidth and probability of packet loss in software-defined WAN. In: 2019 IEEE 15th international conference on the experience of designing and application of CAD systems (CADSM) proceedings. IEEE, pp 1–5. https://doi.org/10.1109/CADSM.2019.8779321
5. Odom W (2019) CCNA 200-301 official cert guide, vol 2. Cisco Press
6. Strelkovskaya I, Solovskaya I, Grygoryeva T, Paskalenko S (2016) The solution to the problem of the QoS characteristics definition for self-similar traffic serviced by the W/M/1 QS. In: 2016 third international scientific-practical conference problems of infocommunications science and technology (PIC S&T) proceedings. IEEE, pp 40–42. https://doi.org/10.1109/INFOCOMMST.2016.7905330
7. Strelkovskaya I, Solovskaya I, Makoganiuk A (2017) Optimization of QoS chracteristics of self-similar traffic. In: 2017 4th international scientific-practical conference problems of infocommunications. science and technology (PIC S&T) proceedings. IEEE, pp 497–500. https://doi.org/10.1109/INFOCOMMST.2017.8246447
8. Strelkovskaya I, Solovskaya I, Makoganiuk A (2021) Different extrapolation methods in problems of forecasting. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in information and communication technology and systems, MCT 2019. Lecture notes in networks and systems, vol 152. Springer, Cham, pp 217–228. https://doi.org/10.1007/978-3-030-58359-0_12
9. Romanov O, Mankivskyi V (2019) Optimal traffic distribution based on the sectoral model of loading network elements. In: 2019 IEEE international scientific-practical conference problems of infocommunications, science and technology (PIC S&T) proceedings. IEEE, pp 683–688. https://doi.org/10.1109/PICST47496.2019.9061296
10. Shahriar F, Fan J (2020) Performance analysis of FHRP in a VLAN network with STP. In: 2020 IEEE 3rd international conference on electronics technology (ICET) proceedings. IEEE, pp 814–818. https://doi.org/10.1109/ICET49382.2020.9119624
11. Anwar U, Teng J, Umair HA, Sikander A (2019) Performance analysis and functionality comparison of FHRP protocols. In: 2019 IEEE 11th international conference on communication software and networks (ICCSN) proceedings. IEEE, pp 111–115. https://doi.org/10.1109/ICCSN.2019.8905333
12. First hop redundancy protocols configuration guide. Cisco IOS Release 15.1 M&T (2018)
13. Hinden R (2004) Virtual router redundancy protocol (VRRP). RFC 3768, Internet Engineering Task Force (IETF)

14. Nadas S (2010) Virtual router redundancy protocol (VRRP) version 3 for IPv4 and IPv6. RFC 5798, Internet Engineering Task Force (IETF)
15. Koch F, Hansen KT (2006) Redundancy performance of virtual network solutions. In: 2006 IEEE conference on emerging technologies and factory automation proceedings. IEEE, pp 328–332. https://doi.org/10.1109/ETFA.2006.355395
16. Nur R, Saharuna Z, Irmawati I, Irawan I, Wahyuni R (2018) Gateway redundancy using common address redundancy protocol (CARP). Int J Inf Technol Electr Eng (IJITEE) 2(3):71–77. https://doi.org/10.22146/ijitee.43701
17. Denis F (2013) UCARP. https://github.com/jedisct1/UCarp
18. Danhieux P (2004) CARP-the free fail-over protocol. GSEC, Practical v1.4b, SANS Institute, pp 1–16
19. Juniper Networks ScreenOS Release Notes. Juniper Networks, Inc. (2018)
20. Rahman ZU, Mukhtar S, Khan S, Khan R, Ullah Z, Rashid R, Ahmad W (2017) Performance evaluation of first HOP redundancy protocols (HSRP, VRRP & GLBP). J Appl Environ Biol Sci 7(3):268–278
21. Sujitha S, Priya KP, Pragathi B (2016) Fault tolerant SDN controller: a survey. Int J Adv Res 186–191
22. Das RK, Pohrmen FH, Maji AK, Saha G (2020) FT-SDN: a fault-tolerant distributed architecture for software defined network. Wireless Pers Commun 114(2):1045–1066. https://doi.org/10.1007/s11277-020-07407-x
23. Hyojoon K, Schlansker M, Santos JR, Tourrilhes J, Turner Y, Feamster N (2012) CORONET: fault tolerance for software defined networks. In: 2012 20th IEEE international conference on network protocols (ICNP) proceedings. IEEE, pp 1–2. https://doi.org/10.1109/ICNP.2012.6459938
24. Li Q, Liu Y, Zhu Z, Li H, Jiang Y (2019) BOND: flexible failure recovery in software defined networks. Comput Netw 149:1–12. https://doi.org/10.1016/j.comnet.2018.11.020
25. Hu T, Yi P, Lan J, Hu Y, Sun P (2020) FTLink: efficient and flexible link fault tolerance scheme for data plane in Software-Defined Networking. Futur Gener Comput Syst 111:381–400. https://doi.org/10.1016/j.future.2019.11.015
26. ElDefrawy K, Kaczmarek T (2016) Byzantine fault tolerant software-defined networking (SDN) controllers. In: 2016 IEEE 40th annual computer software and applications conference (COMPSAC) proceedings. IEEE, pp 208–213. https://doi.org/10.1109/COMPSAC.2016.76
27. Sidki L, Ben-Shimol Y, Sadovski A (2016) Fault tolerant mechanisms for SDN controllers. In: 2016 IEEE conference on network function virtualization and software defined networks (NFV-SDN) proceedings. IEEE, pp 173–178. https://doi.org/10.1109/NFV-SDN.2016.7919494
28. Lemeshko O, Mersni A, Yeremenko O, Omowumi SO, Volotka V, Al-Dulaimi AMK (2021) Application prospects of first hop redundancy protocols for fault-tolerant SDN controllers: a survey. In: 2021 IEEE international conference on problems of infocommunications. Science and technology (PIC S&T) proceedings. IEEE, pp 1–5
29. Lemeshko O, Nevzorova O, Hailan A (2014) The increasing convergence of coordination procedure in the implementation of multipath hierarchical routing. In: 2014 first international scientific-practical conference problems of infocommunications science and technology proceedings. IEEE, pp 45–48. https://doi.org/10.1109/INFOCOMMST.2014.6992294
30. Lemeshko O, Yeremenko O, Mersni A (2021) Investigation of enhanced mathematical model for traffic engineering fault-tolerant routing. In: 2021 international conference on engineering and emerging technologies (ICEET) proceedings. IEEE, pp 1–6. https://doi.org/10.1109/ICEET53442.2021.9659606
31. Lemeshko O, Yeremenko O, Mersni A, Yevdokymenko M (2021) Resilience aware traffic engineering FHRP solution (Invited Paper). In: 2021 IEEE international conference on information and telecommunication technologies and radio electronics proceedings. IEEE, pp 88–92

# Research of Automated Control Systems Development Based on "Publish-Subscribe" Technology Over Low-Bandwidth Radio Networks




**Irina Strelkovskaya and Roman Zolotukhin**

**Abstract** The development of automated control systems (ACS) based on low-bandwidth communication networks of UHF/VHF radio stations is the major problem concerning the system nodes intercommunication and information exchange. The high-throughput communication networks have solved this problem by means of algorithms and protocols based on "publish-subscribe" technology. However, it is necessary to research the ability to use such technology over low-bandwidth communication networks. This research is focused on automated control systems development based on Message Queuing Telemetry Transport (MQTT) and Data Distribution Service (DDS) by Object Management Group (OMG) protocols. The ability to use such mechanism of data exchange over governmental ACS of the low echelon management level based on VHF/UHF radio stations with high requirements to safety, liveliness and reliability are shown. The MQTT, MQTT-SN and DDS protocols analysis was performed. The data about service packets quantity, lost packets quantity, maximal packet size, actual volume of service and information data was performed and the time of communication establishment between MQTT, MQTT-SN and DDS nodes was defined. The recommendations concerning "publish-subscribe" protocols usage in ACS development were given.

**Keywords** Publish-subscribe · MQTT · MQTT-SN · DDS · ACS · QoS · Traffic characteristics · UHF/VHF radio station · Low-bandwidth network · Broker · Gateway · RTPS

I. Strelkovskaya (✉) · R. Zolotukhin (✉)
International Humanitarian University, Rishelievska Street 28, Odesa 65000, Ukraine
e-mail: i.strelkovskaya@mgu.edu.ua

R. Zolotukhin
e-mail: zolotukhinrv@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_13

## 1 Introduction

The relevant problem of Automated Control Systems (ACS) is the search of solutions about information exchange and intercommunication between nodes of the system [1–4]. This research is focused on ACS of low echelon management level with constant users relocation. Such systems commonly operate over communication networks based on UHF/VHF radio stations [1, 5]. That is why data transmission technologies should meet the following requirements [1]:

- increased reliability;
- operational system scalability;
- the ability to function over low-bandwidth communication networks;
- safety and liveliness.

Internet of Things (IoT) netwoks widely use "publish-subscribe" model of algorithms and protocols for data transmission [6, 7]. The "publish-subscribe" model is a design pattern for message transmission. This mechanism states that the nodes of information system can play two roles: "publisher", which generates messages and "subscriber", which receives messages. The system node can play both of these roles simultaneously. As soon as "publisher" generates a message dedicated to a specific "topic" all "subscribers" that listen to that "topic" receive the corresponding message. Therefore, every "topic" creates a list of data receivers. This way "publish-subscriber" pattern implements asynchronous information transmission method, where the message sender is detached from message receiver. At the same time, subscribers are connected to each other on information packets level only, which allows them to be located on different physical nodes and application platforms. Moreover, the usage of mailing lists allows to organize different QoS levels.

The most widely used protocols based on "publish-subscribe" pattern are Message Queuing Telemetry Transport (MQTT) [8–14] and Data Distribution Service (DDS) [15–18]. MQTT is a lightweight open-source and simple protocol, which implements three QoS levels and is used mainly for "machine-machine" interaction and IoT networks [9, 17, 18]. On the other hand, DDS is more complex protocol that has the NATO standard status and is widely used for development of governmental automated control systems [19].

MQTT and DDS implementations assume their usage over low-bandwith communication networks. However, it is necessary to research the possibility of using such mechanisms for development of ACS for low echelon management level with constant users relocation based on UHF/VHF radio stations while such communication links have low throughput and significant data transmission delay, big data transmission delay jitter, high data loss probability etc. [5].

## 2 "Publish-Subscribe" Technologies Analysis

### *2.1 MQTT Protocol Analysis*

MQTT protocol was originally designed for the purpose of information exchange over low-bandwidth networks that have no predetermined transmission delay and unstable communication channel. Simplicity and lightness of this protocol has allowed it to be used for information transmission between devices with limited computational resources and time of autonomous operation.

The MQTT implementation offers a centralized approach for network topology design. Communication system based on MQTT consists of "publisher", "broker" and "subscriber". "Publisher" does not require any data about "subscribers" concerning their location or quantity, it just sends messages by topics. "Subscribers" also do not need any information about "publishers", they just make "broker" aware of topics that they want to track. "Broker" itself is a central MQTT node that ensures interaction between "publishers" and "subscribers". The data exchange between clients is implemented only via broker. "Broker" is responsible for data reception from clients, data processing, data preservation and data delivery control. The quantity of "publishers", "subscribers" and "brokers" on the network is not limited. However, it is worth mentioning that there should be one main "broker" anyway. MQTT architecture is shown in the Fig. 1 [8].

MQTT protocol functions on 7-th level of the OSI model and uses TCP/IP protocol (4-th level) as transport. However, MQTT for Sensor Networks (MQTT-SN) uses UDP/IP as transport, which allows to use both multicast and unicast message delivery. MQTT-SN is designed in the same way as MQTT, but adopted to the needs of wireless communication networks, for example, low throughput, high probability of

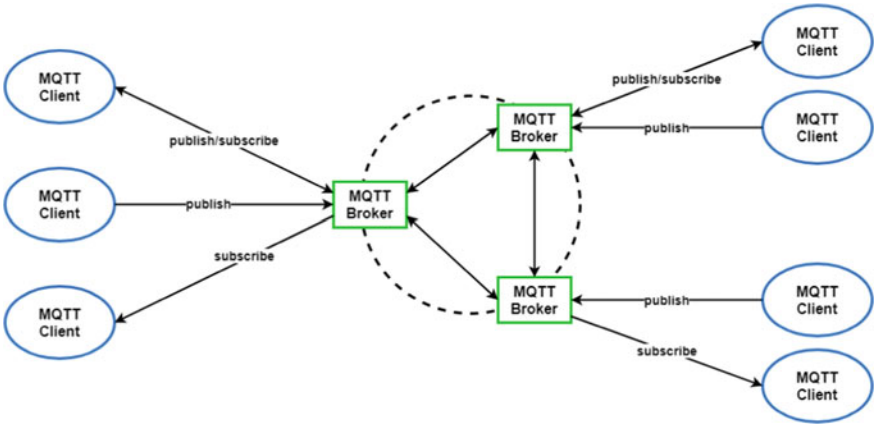



**Fig. 1** MQTT protocol architecture

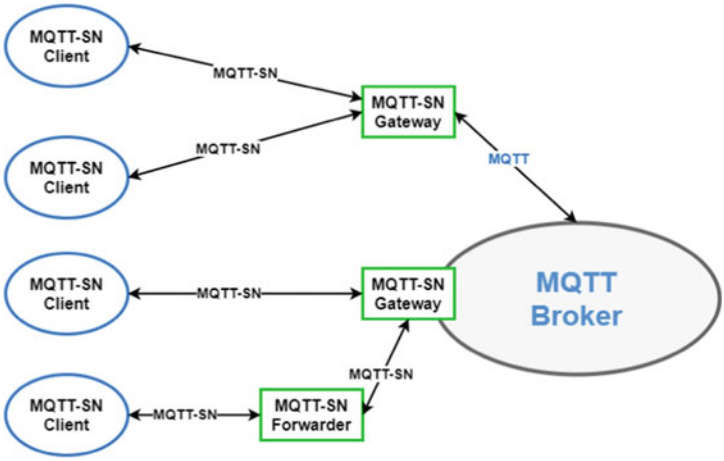



**Fig. 2** MQTT-SN protocol architecture

link break, short message length. MQTT-SN protocol architecture is shown in the Fig. 2 [20].

MQTT-SN protocol describes MQTT-SN client, MQTT-SN gateway and MQTT-SN forwarder. The main MQTT "broker" also plays the role of central node. MQTT-SN client receives or sends messages and connects to the MQTT broker via the MQTT-SN gateway by means of the MQTT-SN protocol. The main function of MQTT-SN gateway is the translation of MQTT-SN packets into MQTT form and vice versa. Moreover, the gateway is able to concatenate several MQTT-SN packets and forward the obtained structure to MQTT broker. The loss of connection between gateway and MQTT broker leads to data transmission between local clients without MQTT broker. MQTT-SN forwarder encapsulates packets received from clients and forwards them to gateway without any change, and decapsulates packets received from gateway and sends them to clients without any change.

MQTT and MQTT-SN protocols support three QoS levels [9, 20]:

- level 0: At most once delivery;
- level 1: At least once delivery;
- level 2: Exactly once delivery.

"Publisher" sends messages with QoS 0 level by default. This means that the message is published on broker but does not require the guaranteed delivery to "subscribers". QoS level 1 means that publisher keeps on publishing the message again and again until it receives the acknowledgement. This way "subscriber" receives the message at least once. QoS level 2 uses additional service messages of acknowledgement and publication finish which ensures the guaranteed message delivery without the possible information duplication. The message is published without any changes at all QoS levels, only header is appended. The header contains information about topic name, topic length, total packet length, QoS level, etc. MQTT-SN protocol

supports one more QoS level -1, when the client publishes message for a gateway without acknowledgement that the gateway is available. What is more MQTT-SN clients also support multicast addressing in order to discover local gateways, however data transmission is still performed by unicast addressing.

The MQTT protocol supports user authentication and provides special fields for sending username and password in the service message CONNECT [8, 10, 11]. The combination of username and password is transmitted in clear text, but MQTT allows the additional use of Transport Layer Security (TLS) and Secure Sockets Layer (SSL) to encrypt all data transmitted over the network [10, 11]. The current MQTT-SN specification [13, 20] doesn't provide user authentication, and only the client ID can be transferred to the broker.

One of the important problems in building the governmental ACS of the low echelon management level is the implementation of mechanisms to control user access to information and system, but MQTT and MQT-SN don't support any algorithms and mechanisms for access control [10, 11]. Some implementations of the MQTT broker, such as Mosquitto, allow authorize users and create an Access Control List (ACL) to differentiate users' access to publications and subscriptions on various topics.

Data integrity verification is not provided in the MQTT and MQTT-SN protocols according to specifications [8, 12, 20]. These functions, if necessary, should be implemented by software that uses these protocols. Thus, the MQTT and MQTT-SN protocols are simple protocols for sending messages under the "publish-subscribe" template, in which the mechanisms and algorithms of information protection are not fully implemented. However, the governmental ACS of the low echelon management level have increased requirements for system security, so it is necessary to analyze another implementation of the "publish-subcribe" template—the DDS protocol.

## *2.2 DDS Protocol Analysis*

DDS protocol is the protocol for information transmission by means of "publish-subscribe" pattern, which is data oriented and designed for highly dynamic distributed systems. Object Management Group (OMG) has developed a standard for it. The data is published within DDS domain and clients can subscribe to receive this data without the knowledge about its' origin or structure, due to the fact that packet already describes it. Clients can publish and receive messages only within their own domain and have their own authorization certificates. DDS is one of the protocols used in such industries as railway networks, aircraft traffic management, smart energy generation, medical care, military and aerospace industry and industrial automation [16–19].

DDS uses transport framework for data transmission, which supports several transmission protocols: TCP, UDP unicast, UDP multicast, shared memory and Realtime Publish-Subscribe (RTPS) wire protocol. Unlike MQTT, DDS supports both the centralized and decentralized topology of communication network.

Centralized approach is built on the usage of "Information Repository" which is a server that contains information about domain clients and does not take part in data transmission. Intercommunication principle of applications with the use of centralized approach is shown on Fig. 3.

This way "Information Repository" receives topic names that "publishers" publish and "subscribers" receive. The data transmission is performed directly between "publisher" and "subscriber". It should be noted that a DDS client can act as the "publisher" and "subscriber" simultaneously.

Decentralized approach state that "subscribers" and "publishers" discover each other by means of RTPS protocol (Fig. 4). It means that every "publisher" discovers its' subscribers itself and stores information about them.

Messages transmission is performed by means of TCP and UDP network protocols and even based on shared memory. Unlike MQTT, DDS is able to transmit data using both unicast and multicast addressing. Moreover, the users quantity is not limited. The development of DDS domain starts with data structure for every topic using Interface Definition Language (IDL). Example of IDL is shown in the Fig. 5.

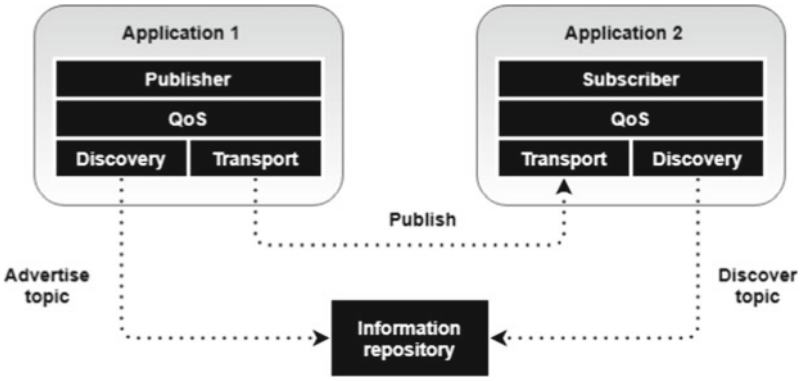



**Fig. 3** Centralized interaction of applications with DDS

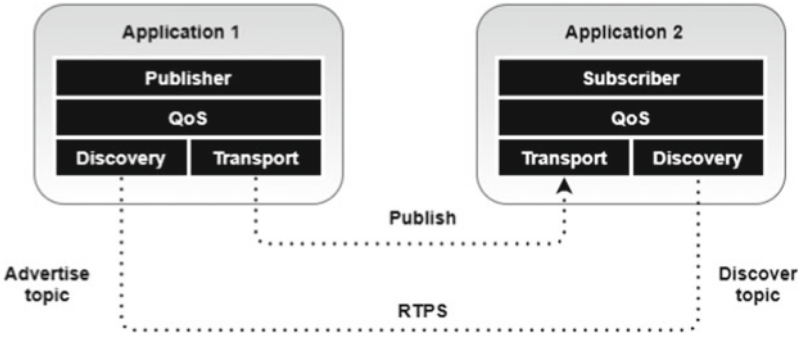



**Fig. 4** Decentralized interaction between applications with DDS

```
module Messenger {
  @topic
  struct Message {
    string from;
    string subject;
    @key long subject_id;
    string text;
    long count;
  };
};
```

**Fig. 5** Part of application code of DDS protocol

DDS protocol implements 22 possible QoS policies. Data policies can be assigned for topic, "subscriber" and "publisher". QoS policies are categorized according to the following criteria:

- time interval between service messages;
- limits of available resources for application operation;
- necessity of and waiting time for reception acknowledgement;
- time for message sending;
- priorities of transport protocols types;
- maximum delivery delay;
- reliability of messages delivery;
- maximum quantity of messages available for retransmission;
- the necessity to store messages.

Let us consider Reliability policy that is responsible for guaranteed message delivery. It include the next two options:

- Best Effort Reliability QoS—the queries for retransmission are absent which means that messages are sent only once;
- Reliable Reliability QoS—this means that can be queries for messages retransmission.

This way, the usage of different QoS policies, different data transmission protocols and different communication system topology within DDS allows to manipulate a set of parameters in order to build ACS. Meanwhile the implementation of DDS protocol is much more complex then MQTT has.

The DDS protocol, in contrast to the MQTT and MQTT-SN protocols, supports the Security Model to implement mechanisms of protecting and encrypting information, user authorization and delimitation of access rights [15, 16]. The Security Model defines the security principles of system users, the objects to be protected, and the operations with objects to be restricted.

The Security Model of the DDS protocol implements the following functions [15]:

- privacy of data;
- data and message integrity;
- authentication of subscribers and publishers;
- authorization of subscribers and publishers;

- message-origin authentication;
- data-origin authentication.

The Service Plugin Interface is provided in the DDS protocol to implement the above Security Model functions. The following plugins can be connected using this interface:

- authentication service plugin that provides the means to verify the identity of the program and / or user and the means to perform mutual authentication between system users and establish a common encryption key based on RSA algorithms;
- access control service plugin that provides the means to enforce access rights policy decisions on transactions that can be performed by authenticated users;
- cryptographic services plugin that implements all cryptographic operations based on RSA, AES, MD5, SHA algorithms, including encryption, decryption, hashing, digital signatures, and includes key distribution tools in the DDS domain;
- logging service plugin that provides auditing of all events related to the DDS security model;
- data tagging service plugin, which provides the ability to tag data and messages for processing according to the QoS policies used.

## 3 Initial Data for Research "Publisher-Subscriber" Model in ACS

The diagram shown in the Fig. 6 is applied to research the possibility of using protocols "Publisher-Subscriber" in governmental ACS of the low echelon management level with constant users relocation. Three Local Computing Networks (LAN) interact with each other through the UHF/VHF radio stations. The radio stations RF-7850M-HH manufactured by Harris are used there. The power of the radio stations is set to 1 W, the operation mode is FF–a narrowband mode with a fixed carrier frequency. Three personal computers (PCs), connected to each other through the Ethernet switch, are installed in every LAN. The switches are connected to radio stations with a special Ethernet cable with RJ-45 (12,067-5220-01) from accessories kit for RF-7850M-HH.

According to the work [5], we will carry out measurements of bandwidth, medium ping and jitter of low-bandwidth radio network and get the following parameters:

- bandwidth—102 kbps;
- jitter—126 ms;
- data loss in channel—2%;
- medium ping—1189 ms.

Ubuntu 20.04.1 LTS operating system is installed on PC. The following software for the implementation of the "Publisher-Subscriber" model is installed on the PC as well:

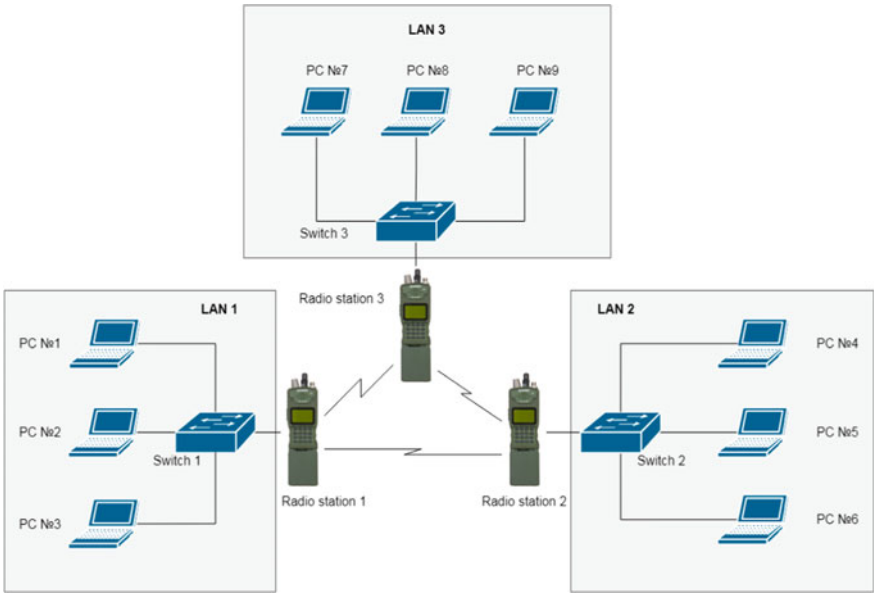



**Fig. 6** Diagram of communication network for measurement

- Mosquitto Utility 2.0.10, which implements the "publisher", the "subscriber" and a broker of the MQTT protocol;
- RSMB utility that has been compiled from source code that implements MQTT-SN gateway;
- software scripts in the python programming language that implements MQTT-SN "Publisher" and "Subscriber";
- software is developed with OpenDDS software source codes in C++ programming language for implement DDS and RTSP protocols;
- "Wireshark" software is used as a traffic analyzer for computer communication networks.

It is necessary to analyze the number of service packets, the number of lost packets, the maximum size of information packets, the actual size of information and service data as well as to determine the time of communication establishment between the nodes of the system in order to research the possibility of using the "Publisher-Subscriber" model in governmental ACS of the low echelon management level with constant users relocation. Let us consider the transfer of a message for basic situations, respectively, dissemination of information in the ACS:

- the transmission of a message when a connection is establishing between the nodes of the system;
- the transmission of messages of different sizes for evaluation the maximum permissible message size;

- the transmission of a message from one client to another, that is from the "publisher" to the "subscriber";
- the transmission of a message from the client to all customers, that is, from the "publisher" to all "subscribers".

## 4 Results of the Research the "Publisher-Subsector" Model

### *4.1 Research of MQTT, MQTT-SN, DDS Protocols Traffic Parameters*

The diagram is being built, according to the Fig. 6 to analyze the number of service packets, the number of lost packets, the maximum size of information packets, the actual size of information and service data and to determine the time to establish a connection between the nodes of the MQTT, MQTT-SN, DDS protocols system. There are MQTT brokers and MQTT-SN gateways configured on the PC No. 1, PC No. 4 and PC No. 7. PC No. 1 is the main broker for MQTT and MQTT-SN protocols. To implement broker of MQTT protocol we used Mosquitto utility, version 2.0.10. To implement gateway of MQTT-SN broker we compiled RSMB utility from source code. There are "publishers" and "subscribers" configured on all PCs. To impement clients of MQTT protocol we used client side Mosquitto utility, version 2.0.10 and to implement clients of MQTT-SN protocol we used software scripts in the python programming language from RSMB utility. We will use QoS 0, 1, 2 levels to research of MQTT and MQTT SN protocols' traffic parameters. Software implementations of the MQTT and MQTT-SN protocols will not work at the same time, but will be turned on alternately, depending on the experiment. After each experiment, all nodes of the system will be restarted to clear the residual data. Wireshark software runs on every PC during MQTT and MQTT-SN traffic characteristics experiments. The diagram of MQTT and MQTT-SN elements shown on Fig. 7.

The software has been created by us using the C++ programming language and OpenDDS source codes for implementation publishers and subscribers of the DDS protocol. The software is created with the support of TCP and UDP transport protocols, moreover, UDP with supporting multicast addressing. A decentralized interaction of applied programs is implemented. To find out "publishers" and their "subscribers" we uses the RTPS protocol. There are also "publishers" and "subscribers" configured on all PCs. Brokers or gateways are not used in this protocol [15]. We will use Best Effort Reliability and Reliable Reliability QoS policy that is responsible for guaranteed delivery of messages for research DDS traffic parameters. Software implementations of the MQTT, MQTT-SN and DDS protocols will not work at the same time. Only one transport protocol UDP or TCP of the DDS protocol will work at a time. RTPS will work throughout all experiments with DDS, as without it publishers won't be able to send messages to subscribers. After each experiment, all

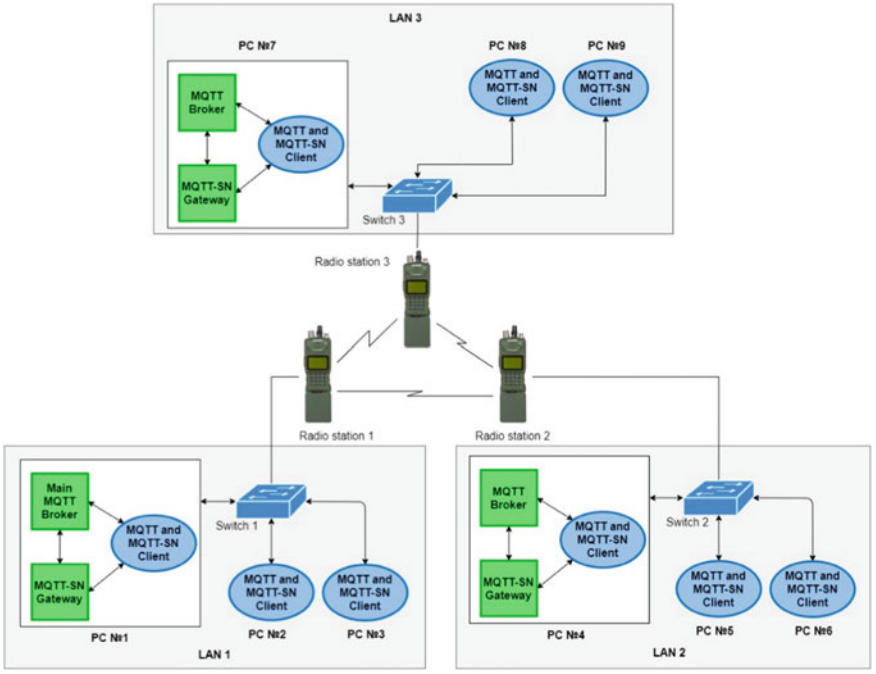



**Fig. 7** Diagram of communication network with MQTT and MQTT-SN protocols

nodes of the DDS system will be restarted to clear the residual data. Wireshark software runs on every PC during DDS traffic characteristics experiments. The diagram of communication network with DSS elements shown on Fig. 8.

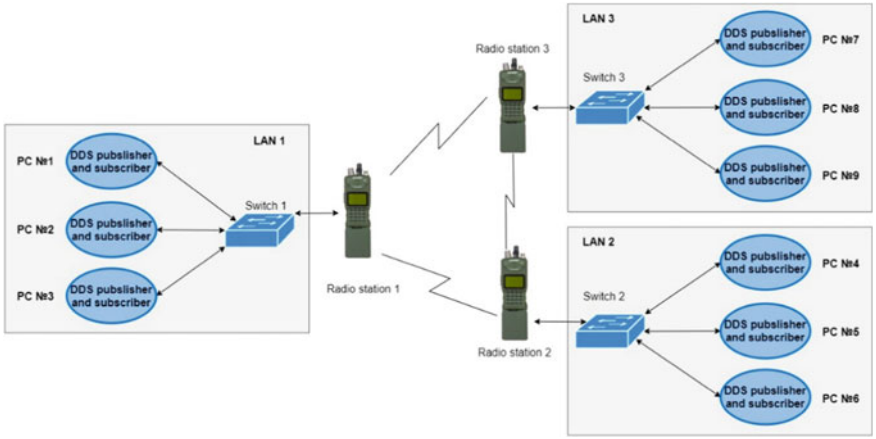



**Fig. 8** Diagram of communication network with DDS protocol

**Table 1** Traffic statistics when MQTT and MQTT-SN protocols establish connection

| QoS level | Traffic parameter | Average value for MQTT | Average value for MQTT SN |
|---|---|---|---|
| 0 | The number of packets | 12 | 15 |
| | The number of service packets | 11 | 14 |
| | Information message time receiving | 6,5 s | 6,4 s |
| 1 | The number of packets | 13 | 16 |
| | The number of service packets | 12 | 15 |
| | Information message time receiving | 6,8 s | 6,7 s |
| 2 | The number of packets | 13 | 16 |
| | The number of service packets | 12 | 15 |
| | Information message time receiving | 6,7 s | 6,5 s |

First experiment: configure MQTT protocol to work accordingly the QoS level 0. Turn off all brokers, "subscribers" and "publishers". Turn on all brokers, "publishers" and "subscribers" and, at the same time, send a message from PC No. 3 to PC No. 9. Using Wireshark software, we repeat these actions 10 times to get the packets dump. Conduct this experiment for MQTT with QoS levels 1 and 2. Repeat the same actions as we have done for the MQTT protocol, but with the MQTT-SN protocol. Table 1 shows the statistics for the traffic transmitted to a network when establishing a connection between the nodes of the system when using MQTT, MQTT-SN protocols.

Clearly that the number of MQTT protocol packets exceeds the number of information packets that is why there is necessity to install TCP and MQTT connection between the nodes of the system. In addition, at the QoS levels 1 and 2 the number of service packets exceeds by 1 compared to the level 0, what is explained by the acknowledgment of receiving for each information message. However, the time of receipt of the information message only depends on the telecommunication network. The number of observed MQTT-SN service packets is more compared to the case of using the MQTT protocol. This can be explained through the necessity in additional MQTT-SN service messages. Such messages are sent as multicast and are designed to identify the available gateways, "subscribers", "publishers".

We turn off brokers, gateways, "publishers" and "subscribers" of MQTT and MQTT-SN protocols; configure DDS protocols according to the Best Effort QoS Reliability policy. We turn on all "publishers" and "subscribers" and at the same time send a message from PC No. 1 to PC No. 7. Then we repeat these actions 10 times. Conduct

**Table 2** Traffic statistics when DDS protocol establishes connection

| QoS policy | Traffic parameter | Average value |
|---|---|---|
| Best effort | The number of packets | 195 |
| | The number of service packets | 194 |
| | Information message receiving time | 43,5 s |
| Reliable | The number of packets | 196 |
| | The number of service packets | 195 |
| | Information message receiving time | 44,8 s |

this experiment for DDS with QoS Reliable reliability policy. Table 2 shows the statistics for the traffic transmitted to a network when establishing a connection between the nodes of the system when using DDS protocol.

The number of service packets and the time of receiving for the DDS protocol is significantly exceed the same parameters of the MQTT and MQTT-SN protocols, because there is a necessity to establish a connection between all nodes of the system. The number of service packets exceeds by one in the QoS Reliable reliability policy compared to the Best Effort policy, which is explained by acknowledgment the receipt of the information message.

Second experiment: we configure MQTT protocol to work accordingly the QoS 0 level, turn on all brokers, "subscribers" and "publishers". Then we begin to transfer information messages of the 100-byte size from PC No. 1 to PC No. 4 after establishing a connection between all nodes of the system, and after each successful transfer, increase the size of the message by 1 byte, moreover the "topics" in information messages occupies 5 bytes. We conduct this experiment for MQTT protocol with QoS levels 1 and 2. The experimental results show that the maximum size of the information packet is not limited, regardless of the QoS level, because the packets, whith the length exceeding MTU (1500 bytes), are fragmented by the TCP/IP protocols into fragments and are sent as parts that are reconstructed by the recipient. We repeat the same actions for the MQTT-SN protocol. The maximum delivered message is 1472 bytes for QoS level 0 and 1470 bytes for levels 1 and 2, while the length of the "topic" is 6 bytes. The reason for this is that MQTT-SN is based on the UDP protocol that can't divide messages into fragments to be recover at the receiving side.

DDS protocol is configured in accordance with the "Best Effort Reliablity" QoS policy and use TCP protocol as transport. The data transmission of packets whithe length of 100 bytes, is performed from PC No. 1 to PC No. 7 after successful connection between all nodes of the system and after every successful transmission we increase packet length by one byte. This experiment is repeated, using UDP protocol. The same experiment is performed with the "Reliable reliability" QoS policy. The experimental results show that the maximum size of the information packet is not limited, regardless of the QoS policy using TCP protocol. The maximum delivered message using UDP protocol is 1448 bytes. The maximum length of the DDS message is 20 bytes less than in the message of the MQTT-SN protocol. The reason for this is the header of the DDS information package has more service information.

Third experiment: we configure MQTT protocol to work accordingly QoS level 0. The transmission of 1000 information messages with the length of 300 bytes is performed from PC No. 2 to PC No. 5 every 10 s, after successful connection between all nodes of the system. At first, information message from PC No. 2 is transmitted to a local broker on a PC No. 1, from PC No. 1 message is transmitted to a remote broker on PC No. 4. At the end, the message is delivered to the final "subscriber" of PC No. 5 from the remote broker on the PC No. 4. The experiment is performed with the QoS levels 1 and 2. The message transmission path remained unchanged, but for each message was additionally sent acknowledgment of receiving the message. The actual volume of transmitted useful information is 300 Kbytes. The same actions were performed with the MQTT-SN protocol. The statistics of traffic transmitted in the radio channel by MQTT and MQTT-SN protocols are shown in the Table 3.

At the QoS level 0, due to losses in the radio channel, 18 information messages of the MQTT and 12 MQTT-SN messages were not delivered. In addition, 200 special service packets were sent to check the availability of brokers. At QoS levels 1 and 2 there were no lost messages, but the number of service messages significantly increased due to the necessity to confirm of receiving of each information message and re-request of lost data. The MQTT-SN protocol has transferred 1000 service messages more, compared to the MQTT, due to additional MQTT-SN service

**Table 3** The traffic statistics of sending messages by MQTT and MQTT-SN protocols

| QoS level | Traffic parameter | Average value for MQTT | Average value for MQTT SN |
|---|---|---|---|
| 0 | The total number of sending packets | 1200 | 2200 |
| | The number of service packets | 200 | 1200 |
| | The number of lost packets | 18 | 12 |
| | The volume of MQTT data transmitted | 312,4 Kbytes | 317,4 Kbytes |
| 1 | The total number of packets sent | 2239 | 3246 |
| | The number of service packets | 1226 | 2226 |
| | The number of lost packets | 0 | 0 |
| | The volume of MQTT data transmitted | 322,5 Kbytes | 323,7 Kbytes |
| 2 | The total number of packets sent | 2236 | 3244 |
| | The number of service packets | 1224 | 2225 |
| | The number of lost packets | 0 | 0 |
| | The volume of MQTT data transmitted | 322,1 Kbytes | 323,2 Kbytes |

messages that are sent as multicast. Similarly, 200 special service packets sent by gateways in order to check the availability of the main broker. At the QoS levels 1 and 2, no messages have also been lost.

DDS protocol is configured in accordance with the "Best Effort Reliablity" QoS policy and uses UDP protocol as the transport. The transmission of 1000 information messages with the length of 300 bytes is performed from the PC No. 2 to PC No. 5 every 10 s, after successful connection between all nodes of the system. Then we set up DDS for working with TCP protocol, change DDS settings for working with Reliable Reliablity QoS policy and repeat the experiment. All information messages from the PC No. 1 are immediately transferred to PC No. 7. The actual volume of transmitted useful information is 300 Kbytes. 13 messages were not delivered during Best Effort QoS policy using UDP protocol due to loss in the radio channel. In other cases, there were no message losses. However, compared to the MQTT or MQTT-SN, the DDS protocol has much bigger volume of total data and service data transmitted. This is caused by the fact that DDS constantly checks availability of every "subscriber" and "publisher" and service packets are transmitted with the length from 50 to 600 bytes. The statistics for traffic transmitted in radio channel with the DDS protocol are shown in the Table 4.

The fourth experiment: the MQTT protocol is configured to operate at the QoS level 0. The messages are transmitted from the PC No. 2 to all other PCs. It is worthwhile to mention that the messages are sent to the local broker located on the PC No. 1 first. The messages are then sent from the PC No. 1 to "subscribers" on PC No. 2 and PC No. 3. In addition, the messages are delivered to remote brokers located on the PC No. 4 and PC No. 7. Remote brokers retransmit data to local "subscribers". In this way, two informational messages were transmitted through the radio channel. The same experiment is repeated for QoS level 1 and 2. The transmission path did not change and four informational messages were transmitted via the radio channel. The additional packets served for acknowledgement of data reception by remote brokers to main broker.

MQTT-SN protocol is configured to operate at the QoS level 0. The messages from the PC No. 3 are transmitted to all other PC's on the network. It can be mentioned that messages are sent from the PC No. 3 to local gateway on the PC No. 2. The messages are then retransmitted to the local subscribers by means of the MQTT-SN protocol. In addition, the messages are delivered to main broker located on the PC No. 1 using the MQTT protocol. The messages are then sent from the PC No. 1 to remote gateways located on the PC No. 4 and PC No. 7 delivering data to their local "subscribers". In this way, two informational messages were transmitted through the radio channel. The same experiment is repeated for the QoS levels 1 and 2. The transmission path was the same. However, four informational messages were transmitted via the radio channel. The additional packets served for acknowledgement of data reception by the remote brokers to main broker.

DDS protocol is configured in accordance with the Best Effort Reliability QoS policy and uses UDP protocol as the transport. The messages from the PC No. 1 are transmitted to all other PC's. It can be mentioned, that the messages are delivered to all "subscribers" by means of only one packet. The experiment is repeated with

**Table 4** The traffic statistics while sending messages with the DDS protocol

| QoS policy | Type of traffic protocol | Traffic parameter | Average value |
|---|---|---|---|
| Best Effort | TCP | The total number of sending packets | 1125 |
| | | The number of service packets | 125 |
| | | The number of lost packets | 0 |
| | | The actual volume of data transmitted | 339,5 Kbytes |
| | UDP | The total number of packets sent | 2336 |
| | | The number of service packets | 1336 |
| | | The number of lost packets | 13 |
| | | The actual volume of data transmitted | 540,3 Kbytes |
| Reliable | TCP | The total number of sending packets | 2236 |
| | | The number of service packets | 1236 |
| | | The number of lost packets | 0 |
| | | The actual volume of data transmitted | 563,9 Kbytes |
| | UDP | The total number of packets sent | 3229 |
| | | The number of service packets | 2215 |
| | | The number of lost packets | 0 |
| | | The actual volume of data transmitted | 712,3 Kbytes |

TCP protocol as the transport protocol. PC No. 1 has sent 8 messages to every "subscriber" on the network, while the radio channel has transmitted 6 packets. DDS protocol configuration has been changed to operate with the Reliable reliability QoS and the same experiment has been performed. The results demonstrate that the radio channel has transmitted 12 messages with the TCP protocol and 7 messages with the UDP protocol. This is explained by the fact that every "subscriber" has sent an acknowledgement packet to confirm the message delivery.

During the configuration of the telecommunication system and during the research, it was found that if the main MQTT broker is not available or there is no connection with it, the message could be sent only within the local network. When reconnecting local brokers to the main one, it is again possible to exchange messages between users of different LANs. During the configuration and study of the MQTT-SN protocol, it was found that when the gateways lost communication

**Table 5** The results of the MQTT, MQTT SN and DDS protocols research

| Parameter | | MQTT | | | MQTT-SN | | | DDS | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 0 | 1 | 2 | Best effort | | Reliable | |
| | | | | | | | | TCP | UDP | TCP | UDP |
| Establish connection | The number of packets | 12 | 13 | 13 | 15 | 16 | 16 | 195 | | 196 | |
| | The number of service packets | 11 | 12 | 12 | 14 | 15 | 15 | 194 | | 195 | |
| | Information message time receiving, s | 6,5 | 6,8 | 6,7 | 6,4 | 6,7 | 6,5 | 43,5 | | 44,8 | |
| Sending 1000 messages | The total number of sending packets | 1200 | 2239 | 2236 | 2200 | 3246 | 3244 | 1125 | 2336 | 2236 | 3229 |
| | The number of service packets | 200 | 1226 | 1224 | 1200 | 2226 | 2225 | 125 | 1336 | 1236 | 2215 |
| | The number of lost packets | 18 | 0 | 0 | 12 | 0 | 0 | 0 | 13 | 0 | 0 |
| | The actual volume of data transmitted, Kbytes | 312,4 | 322,5 | 322,1 | 317,4 | 323,7 | 323,2 | 339,5 | 540,3 | 563,9 | 712,3 |
| Architecture | | Centralized | | | Centralized | | | Centralized and decentralized | | | |
| Count of users | | Unlimited | | | Unlimited | | | Unlimited | | | |
| Authentication | | Present | | | Absent | | | Present | | | |
| Privacy of data | | Absent | | | Absent | | | Present | | | |
| Encryption | | Present | | | Absent | | | Present | | | |
| Data compression | | Absent | | | Absent | | | Present | | | |
| Access control | | Absent | | | Absent | | | Present | | | |

with the main broker, the message could be transmitted between users of different LANs, but without supporting QoS levels 1 and 2.

The results of the analysis and experiments of the research the traffic characteristics of the MQTT, MQTT-SN and DDS protocols are shown in Table 5.

## *4.2 Recommendations for Use of "Publish-Subscribe" Protocols in ACS*

The performed research results allow to give the next recommendations:

1. The ACS based on low-bandwidth radio channels with significant transmission delay and low throughput allow the usage of "publish-subscribe" protocols. The message delivery reliability requires the usage of QoS level 1 and level 2 for

MQTT and MQTT-SN protocols and "Reliable reliability" QoS policy for DDS protocol.
2. DDS protocol ensures necessary reliability for ACS. MQTT and MQTT-SN protocols loose system integrity in case of main broker fault. MQTT protocol can still be used when the main broker node has backup.
3. The MQTT and MQTT-SN protocols should be used to ensure scalability of ACS and bandwidth savings of low-bandwidth communication channels.
4. The use of the DDS protocol with security model and plugins for information, data and user protection will ensure security and safety of governmental ACS of the low echelon management level. ACS design based on MQTT and MQTT-SN protocols demands the usage of brokers with implementation of information protection algorithms and mechanisms.

## 5 Conclusions

1. MQTT, MQTT-SN and DDS protocols were analyzed. Their usage for development of governmental ACS of the low echelon management level with constant users relocation was offered and recommendations about their usage were given.
2. The application software for research of governmental ACS of the low echelon management level with constant users relocation was created based on C++ programming language and DDS protocol.
3. The parameters of service packets quantity, lost packet quantity, maximum information packet length, actual volume of informational and service data were obtained for MQTT, MQTT-SN and DDS protocols. It was defined that the time for connection establishment between nodes is 6,4–6,8 s for the MQTT protocol and 43–45 s for the DDS protocol.
4. It is possible to determine the load of the MQTT, MQTT-SN and DDS protocols on low-bandwidth communication networks based on the results of the research, which will increase the efficiency use of communication channels and reduce the probability of service failures in ACS based on UHF/VHF radio stations and "publish-subscribe" protocols.

## References

1. Strelkovskaya IV, Zolotukhin RV, Makoganiuk AO (2021) Modeling of telecommunication components of automated control systems in low-bandwidth radio networks. In: Vorobiyenko P, Ilchenko M, Strelkovska I (eds) Current Trends in Communication and Information Technologies. IPF 2020. LNNS, vol 212, pp 150–170. Springer, Cham. https://doi.org/10.1007/978-3-030-76343-5_9
2. Strelkovskaya I, Zolotukhin R, Strelkovska J (2021) Comparative analysis of file transfer protocols in low-bandwidth radio networks. In: Proceedings of the 9rd International Conference on Applied Innovations in IT, vol 9, Is 1, Koethen, Germany, pp 27–32, 9 March 2021

3. Bayne JS (2006) A theory of enterprise command and control. In: 2006 IEEE Military Communications Conference, Washington, DC, USA, 23–25 October 2006. https://doi.org/10.1109/MILCOM.2006.302294
4. Strelkovskaya I, Solovskaya I, Makoganiuk A (2017) Optimization of QoS chracteristics of self-similar traffic. In: 2017 4th International Scientific-Practical Conference Proceedings, Problems of Infocommunications Science and Technology, pp 497–500. https://doi.org/10.1109/INFOCOMMST.2017.8246447
5. Strelkovskaya I, Zolotukhin R (2020) Research of low-bandwidth radio networks QoS parameters. Inf Telecommun Sci Int Res J 11(1(20)), January-June. https://doi.org/10.20535/2411-2976.12020.77-81
6. Pahl M-O (2019) Multi-tenant IoT service management towards an IoT app economy. In: HotNSM workshop at the International Symposium on Integrated Network Management (IM), Arlington, VA, USA
7. Pahl M-O, Carle G, Klinker G (2016) Distributed smart space orchestration. In: NOMS 2016–2016 IEEE/IFIP Network Operations and Management Symposium, pp 979–984. IEEE. https://doi.org/10.1109/NOMS.2016.7502936
8. MQTT Version 3.1.1: Edited by Andrew Banks and Rahul Gupta. 29 October 2014, OASIS Standard. http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html. Accessed November 2021
9. Sarafov V (2018) Comparison of IoT data protocol overhead. In: Network Architectures and Services, vol 23. https://www.net.in.tum.de
10. Niruntasukrat A, Issariyapat C, Pongpaibool P, Meesublak K, Aiumsupucgul P, Panya A (2016) Authorization mechanism for MQTT-based internet of things. In: 2016 IEEE International Conference on Communications Workshops, pp 290–295. IEEE. https://doi.org/10.1109/ICCW.2016.7503802
11. Upadhyay Y, Borole A, Dileepan, D (2016) MQTT based secured home automation system. In: 2016 Symposium on Colossal Data Analysis and Networking (CDAN), pp 1–4. IEEE. https://doi.org/10.1109/CDAN.2016.7570945
12. MQTT security fundamentals: MQTT payload encryption (2015). https://www.hivemq.com/blog/mqtt-securityfundamentals-payload-encryption/. Accessed November 2021
13. Park C-S, Nam H-M (2020) Security architecture and protocols for secure MQTT-SN. IEEE 8:226422–226436. https://doi.org/10.1109/ACCESS.2020.3045441
14. Thota P, Kim Y (2016) Implementation and comparison of M2M protocols for internet of things. In: 2016 4th International Conference on Applied Computing and Information Technology/3rd International Conference on Computational Science/Intelligence and Applied Informatics/1st International Conference on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD), pp 43–48. https://doi.org/10.1109/ACIT-CSII-BCD.2016.021
15. OMG Data Distribution Service (DDS), Version 1.4, April 2015. https://www.omg.org/spec/DDS/1.4. Accessed May 2021
16. Pardo-Castellote G (2003) OMG data-distribution service: architectural overview. In: 23rd International Conference on Distributed Computing Systems Workshops, 2003, Proceedings, pp. 200–206. IEEE
17. Aures G, Lübben C (2019) DDS vs. MQTT vs. VSL for IoT. In: Proceedings of the Seminar Innovative Internet Technologies and Mobile Communications (IITM), Summer Semester, Munich, Germany. https://doi.org/10.2313/NET-2019-10-1
18. Profanter S, Tekat A, Dorofeev K, Rickert M, Knoll A (2019) OPC UA versus ROS, DDS, and MQTT: performance evaluation of industry 4.0 protocols. In: 2019 IEEE International Conference on Industrial Technology (ICIT), 13–15 February 2019, Melbourne, VIC, Australia. https://doi.org/10.1109/ICIT.2019.8755050
19. STANAG 4754 Ed: 1, NATO Generic Vehicle Architecture (NGVA) For Land Systems–AEP-4754 EDITION A, 22 February 2018
20. MQTT For Sensor Networks (MQTT-SN) Protocol Specification. Andy Stanford-Clark and Hong Linh Truong, 14 November 2013. https://www.oasis-open.org/committees/download.php/66091/MQTT-SN_spec_v1.2.pdf. Accessed November 2021

# Multipoint Data Transmission Issues in High Bandwidth-Delay Product TCP/IP Networks




**Nikolai Mareev, Dmitry Kachan, and Eduard Siemens**

**Abstract** Modern network hardware is nowadays ready to provide high-speed channels across the countries and between continents providing wide network resources to end-users. In presence of this, algorithms of multipoint data transmission become more and more a bottleneck, utilizing available network resources not optimally, which can be a consequence of the flaws of modern software solutions and data transport protocols. In the meanwhile, several promising solutions for traffic control have been proposed in the last decade, however the area of multipoint high-speed data transmission remains an insufficiently researched field. This work is aimed to observe the main issues of the data transmission in TCP/IP based wide area networks focusing mainly on congestion control algorithms and software issues in point-to-multipoint solutions. In this paper, the main issues of congestion control algorithms have been reviewed and compared in different network delay and loss probability cases. Proposed earlier Bottleneck Queue Level congestion control (BQL) algorithm has been observed in the context of decreasing the negative impact provided by widely used algorithms on the end-user side. The problem of underutilization of existing channels is described and several software methods of increasing the data throughput have been touched. Several issues of point-to-multipoint data transmission and application-layer multicast solution based on RMDT have been discussion. A final performance evaluation and comparison of TCP BBR and RMDT BQL were made with a tuned Linux TCP/IP stack. The evaluation of introduced algorithms and their prototyping were made with Reliable Multi-Destination Transport Protocol (RMDT), a UDP-based, high-speed transport protocol. All tests have been performed in the emulated WAN environment of Future Internet Lab Anhalt (FILA).

N. Mareev (✉) · D. Kachan (✉) · E. Siemens (✉)
Future Internet Lab Anhalt, Anhalt University of Applied Sciences, Bernburger Street 57, 06366 Köthen, Germany
e-mail: nikolai.mareev@hs-anhalt.de

D. Kachan
e-mail: dmitry.kachan@hs-anhalt.de

E. Siemens
e-mail: eduard.siemens@hs-anhalt.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_14

**Keywords** RMDT · TCP · Multipoint · Big data transport · Congestion control · HBDP

## 1 Introduction

The demand for technical progress in data transmission technologies is growing every year. According to Cisco VNI Complete Forecast Highlights [1], the total IP consumer traffic may reach 332.7 EB per month by 2022. It includes different types of traffic and depending on a certain task, requirements for the data transmission can significantly differ. For example, video streaming applications stand for relatively low data rate flows (e.g. UHD 15–18 Mbps, 1080p 8–12 Mbps), but it generates most traffic across the global network. Data transmission in TCP/IP network is controlled by a stack of protocols delivering a set of features for the applications. Most internet traffic consists of point-to-point flows with reliable data transmission services. So-called Elephant data transmission flows, which stands for long-living high-speed data delivery works in a best-effort environment and require high-performance algorithms to use network resources effectively. Modern wireless technologies like Wi-FI 6 and 5G as part of the chain in the TCP data transmission across the wide-area network forming a wireless heterogeneous network with time-varying parameters. It is a big challenge for researchers and engineers to develop a high-efficient software to optimally utilize resources provided by such a network.

A number of fast data transport solutions have been proposed in recent two decades to increase the utilization of the existing channels. Firstly, UDP-based solutions like RBUDP [2] and TSUNAMI [3] showed significant performance benefits against TCP. Later, another UDP-based solution, UDT [4] demonstrated relatively high data transmission performance with congestion control based on AIMD concepts. Ren Y. et al. in [5] made a comparison of several UDP-based transport protocols and demonstrated the demand for more efficient algorithms of rate and congestion control. Later research made by Yu Se-Young et al. [6] shows the performance benefits of parallel data transfer based on TCP or UDP transport protocols like GridFTP [7] and FDT [8]. Nadig D. et al. in [9] have observed the current data transfer tools from the perspective of its software architecture and applied features. In that, he has shown performance benefits of modern data transmission protocols based on the idea of sending the data in parallel streams. Research in the area of high-speed data transmission is under high interest and new solutions are aimed to increase the performance in modern challenging use cases.

This work is provided under a certain context - Bottleneck Queue Level (BQL) congestion control algorithm [10], proposed in the course of CloudBDT and Bit-Booster projects at the Future Internet Laboratory Anhalt (FILA). Mentioned projects aimed at research in the field of fast data transmissions and operate with a Reliable Multi-Destination Transport protocol (RMDT) [11] and applications based on this protocol. These applications are addressed to satisfy the requirements of high-speed

reliable point-to-point and point-to-multipoint data transmission in high bandwidth-delay product TCP/IP networks even in presence of packet losses and delay jitter. Recently, Karpov K. et al. have described in [12] key methods and solutions in the field of data transmission performance enhancements in high-speed wide area networks proposed by FILA. This work is going deeper into the traffic management issues.

The rest of the paper is organized as follows: Next section describes the FILA testbed topology used further in the performance tests. In the third section, the main network congestion mechanics are pointed out and the so called Bufferbloat problem is mentioned. Section 4 observes congestion control algorithms and their main issues. In the fifth section, the problem of underutilization of the high-speed channels with modern hardware is given. The last section stands for the conclusion.

## 2 Testbed Topology

All test and performance evaluations have been performed in 40 GE Laboratory of Future Internet Lab Anhalt (FILA). The testbed network is presented in Fig. 1.

Core elements in the testbed are Netropy 10G, Netropy 10G2, and Netropy 40 G—WAN emulators from Apposite Technologies. These devices allow emulation of various network conditions by setting the properties of the channel and continuously gathering per-second forwarded data stream statistics such as data transfer rate, tx queue buffer load level, packet losses, etc. An important element in the testbed is a network switch: 10 G switches—Extreme Networks Summit x650-24x with 24 10GBASE-X SFP+ interfaces, 488 Gbps maximum aggregated bandwidth and 363

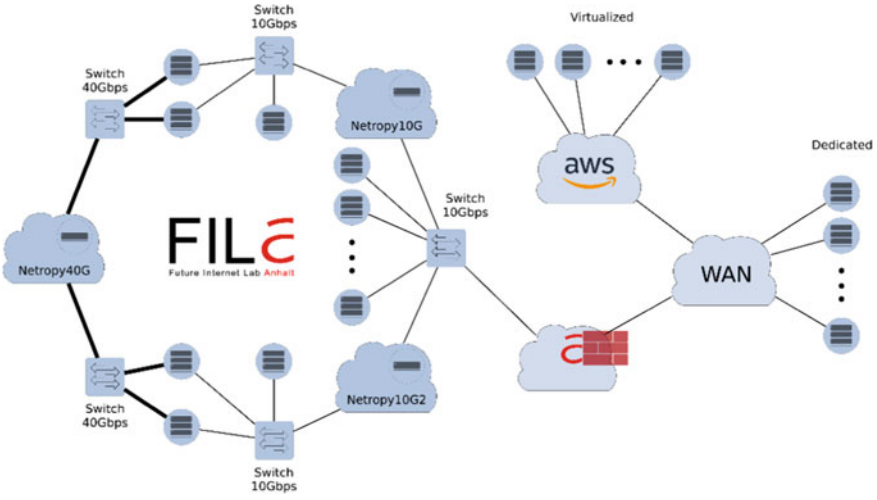



**Fig. 1** FILA test environment

Mpps maximum packet throughput; 40 G switches—Extreme Networks Summit x770-32q with 32 40GBASE-X QSFP+ interfaces, 2560 Gbps maximum aggregated bandwidth and 1.4 Billion packets per second maximum packet throughput. It is an edge-level network switch with tiny shared port queue buffers. The core servers configuration used in the 40 G data transmission consist of Intel(R) Xeon(R) CPU E5-2643 v4 3.40 GHz, 40000baseSR4 full duplex on Emulex Corporation OneConnect NIC, Ubuntu 20.04.3 LTS with GNU/Linux 5.4.0-84-generic x86_64 Kernel and 64 GB DDR4 of RAM. The core servers configuration used in the 10 Gbps data transmission consists of at minimum 32 GB DDR3 of RAM available, Intel(R) Xeon(R) CPU X5690 @ 3.47 GHz, Chelsio Communications Inc T420-CR Unified Wire Ethernet Controller and identical to 40 G core servers OS and kernel. 40 G core servers can also be used in the 10 G data transmission with additional Intel Corporation 82599ES 10-Gigabit SFI/SFP+ NIC.

## 3 Network Congestion Control

A common challenge for most data transmission applications are network congestions. The network congestion occurs when a certain node as part of the network receives more data than it can handle or forward to an output interface. The amount of data that exceeds the node data handling or forwarding capabilities are stored in a FIFO buffer to be transferred delayed what leads to increased network latency for the data transmission routine. The amount of memory in the node buffer is limited and when there is no free space to store a data packet, the new incoming packets will be dropped. This basic queuing management algorithm is called “tail drop”. Network delay and packet drops became general indicators for controlling network congestion. Depending on the certain application, main data transmission performance parameters are varying, however, it is possible to highlight them for a general data transmission how it did Kleinrock L. in [13] - maximizing throughput while minimizing the delay and packets losses caused by congestion. Thus, the main function of the congestion control algorithm is to utilize available bandwidth yet keep network node buffers as free as possible, here and below this will be called an optimal point, as it showed in Fig. 2.

Here, amount data inflight stands for data that has been sent but not yet delivered and/or acknowledged. The acronym RTT stands for Round Trip Time - a time interval between sending the first bit of a data packet to the “wire” by the sending host and receiving the last bit of this packet from the wire at the receiving side. BDP is an acronym that stands for Bandwidth-Delay Product of the certain connection and corresponds to the amount of data inflight in Kleinrock’s optimal point. In TCP terms the amount data inflight corresponds to send window size with the definition “the amount data to send before acknowledge arrives”. Tree colored regions on the figure above stands for different network utilization stages:

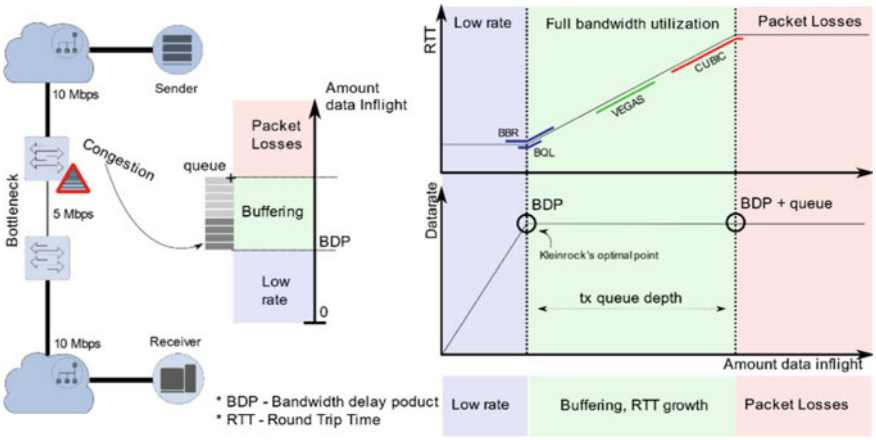



**Fig. 2** Network congestion base mechanics

- 0 < Inflight < BDP: is colored blue and describes the underutilization of the channel. The node buffer is free, there is no additional delay and there are no packet losses.
- BDP < Inflight < BDP + queue: is colored green and describes the situation when the available bandwidth is fully utilized, the node buffer queue is busy what leads to increased network delay.
- BDP + queue < Inflight: is colored red and describes the overloaded bottleneck's queue that causes packet losses and high additional network delay.

Most congestion control algorithms operate in the green-colored region, keeping high channel utilization but yet causing additional network delay and packet losses. Thus, novel congestion control algorithms are needed to utilize network resources more efficiently.

## *3.1 The Bufferbloat Problem*

Special attention is worth to be paid to increased latency in the global network caused by the congestion of excessively large node buffers – the Bufferbloat problem. Such additional network delay can be high enough to significantly decrease the performance of the delay-sensitive applications. In some cases, the network delay can reach a matter of seconds what can lead to the dropping of the existing connections. Additionally, inefficient congestion control algorithms in such deep buffers provoke significant delay jitter, which negatively affects the performance of real-time applications. Deep buffers have been widely included in high-speed routers to satisfy the needs of high-rate data transfers over long pipes. The widely used rule-of-thumb to sizing the router buffers is commonly attributed to a paper by Villamizar C. and Song C. [14].

$$B = RTT \times R \tag{1}$$

This rule implies the requirement for the network routers to have buffers not less than BDP in the link to guarantee full bandwidth utilization with formula 1: (where R is Data Rate and B is buffer depth). Larger network buffers lead to higher network delay during a congestion. The impracticality of the rule-of-thumb in the cases with a high amount of TCP streams was challenged by the Stanford research group in [15]. This work showed that requirements for the buffer depth are decreasing with the increased number of TCP streams in this link. Gettys J. have introduced the Bufferbloat problem to the public in [16]. One possible solution for this issue is Active Queue Management (AQM) described in RFC 7567 [17]. AQM is a link-side solution that allows decreasing the negative effect of the Bufferbloat across the network affecting on buffering process with specific algorithms. Adams R. in [18] did a detailed classification and survey of several proposed AQM algorithms. The first proposed and still common AQM algorithm in use is RED (Random Early Detection) proposed by Floyd S. and Jacobson V. in [19] and its further variants. The main idea is to change the network parameters for a certain link by dropping random packets if the queue load exceeds some threshold. The more gentle way for AQM was proposed in RFC 5562 [20]. Here the Explicit Congestion Notification bit has been proposed to include in TCP packets to indicate the congestion. With such an additional indication the congestion control algorithms can react to the congestion before the network buffer is overloaded. Despite the promising results in decreasing the overall network congestion, this is only a part of the solution. AQM is a link-based solution, which requires configuration of the whole infrastructure. Meanwhile, the main problem is in source-side data transmission protocols and congestion control algorithms which utilize the resources not effectively.

## 4 Congestion Control

A congestion control algorithm (CCA) is a mechanism and a part of the data transmission protocol, aimed at avoiding network congestions by changing the send rate according to the evaluated data transmission parameters. The main principles and definitions of the TCP Congestion Control are described in RFC 5681 [21] by Blanton E. et al. Afanasyev A. et al. did a significant work to systematize and observe the most known congestion control algorithms in [22]. Here he has shown that despite the complexity of these solutions they can be divided into several groups according to the congestion indication they mainly use and its behavior. Some algorithms tend to increase the data rate until the packet losses appear, which leads to a fully occupied network node buffer and significant additional network delay. These algorithms are usually called “loss-based” or “reactive”. Other solutions can evaluate other patterns like specific RTT value (“delay-based”) or receive data rate value. Describing the tend of such solutions not to occupy network node buffer fully, decreasing the negative impact on the network. This group of algorithms is called “proactive”.

### *4.1 Reactive Congestion Control*

Most data transmission flows over the internet nowadays are controlled by loss-based congestion control algorithms and operate close to the fully-loaded network queue if possible (red-colored line in Fig. 2). This group of algorithms includes such notable solutions as TCP NewReno [23], Highspeed TCP [24], BIC [25], CUBIC [26], and similar. TCP NewReno is one of the first such solutions, a default congestion control solution in FreeBSD up to the day. TCP Highspeed was proposed to increase the TCP performance in high-speed networks. BIC was proposed to increase TCP performance and scalability in HBDP links, using a binary Increase function of a congestion avoidance part of the algorithm. TCP CUBIC is a result of further research on the topic of high-speed loss-based algorithms. It is a default congestion control algorithm in the Linux kernel since 2.6.19, in Windows since Windows 10 last updates, and in Windows Server since 2019. And, this is one of the most widely used congestion control on the internet as was shown by Mishra A. et al. in [27]. The typical loss-based flow controlled by CUBIC is presented in Fig. 3. In this test case, the available bandwidth is 8 Gbps, the bottleneck buffer depth is 40 Mbytes, and the RTT is equal to 20 ms. This example shows the tendency of this algorithm to overload the available bottleneck buffer space keeping high channel utilization alongside high network delay and provoking packet drops. Figure 4 demonstrates the evaluation of the average throughput of the TCP CUBIC data transmission in different network use cases.

This performance evaluation was made with WAN emulation done by Netropy WAN Emulator. It is shown that the packet loss probability and higher leads to significant degradation of the channel utilization by the loss-based congestion control algorithm. Thus, packet loss intolerance is one of the issues of modern congestion control algorithms. Sheshadri R. K. and Koutsonikolas D. have evaluated the packet loss rate of 802.11 channels in [28]. The values of probability of packet losses for the emulation are chosen regarding the mentioned research of probabilities of packet losses in a wireless environment. The range of the network delay values for the emulation is chosen to cover most intercontinental use cases.

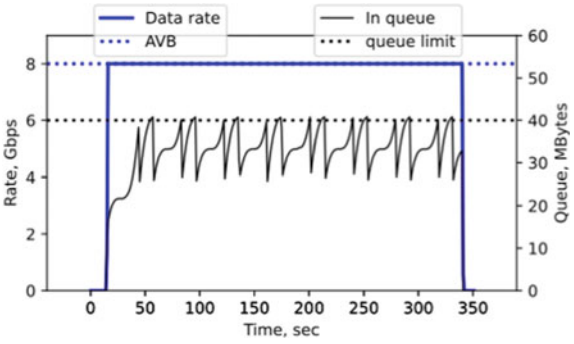



**Fig. 3** TCP CUBIC flow example

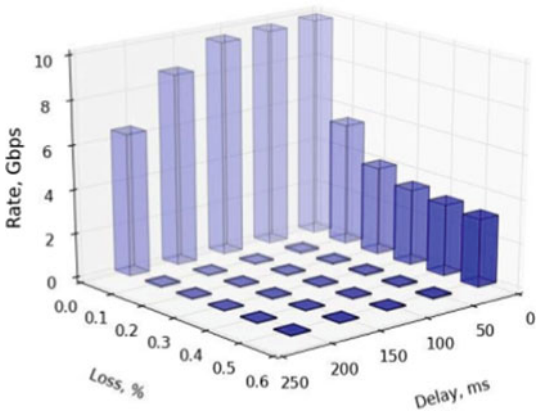



**Fig. 4** TCP CUBIC throughput

## *4.2 Proactive Congestion Control*

One of the first alternative approaches to the loss-based algorithms, TCP VEGAS was proposed by Brakmo, L.S., and Peterson, L.L. in [29]. It is a congestion control algorithm that operates in a proactive way in terms of preventing the bottleneck queue buffer from overloading and congestion-caused packet losses event using delay as the main congestion indication. TCP VEGAS demonstrates possibilities of traffic stabilization and bandwidth utilization keeping the bottleneck queue from being idle. Unfortunately, this and similar approaches demonstrate significant performance degradation during the simultaneous coexistence of delay-based algorithms with any loss-based solution. During further research, FAST [30] became another notable solution in terms of delay-based congestion control algorithms. FAST, introduced by Ren Y. et al. was aimed at solving the main issues of a delay-based CCA while keeping its advantages. This approach showed increased performance and resource sharing, however keeps several problems. FAST evaluates minimal evaluated RTT of the link during data transmission as base link delay for its operating purposes, which makes it problematic for such an approach to work in a dynamically changing environment (e.g. suddenly changing network path). Performance of such an approach decreases in shallow buffer scenarios as is shown in [31] by Jamali S. et al. Also, controlled by TCP FAST, data transmission requires a specific amount of memory in the node queue buffer, and with growing number of such connections amount of occupied memory is growing too. The amount of occupied memory is defined in algorithm parameters and used as a setpoint for the congestion control process. This may result in dependency on the data transmission performance and limits the number of such connections in a single link.

A relatively new congestion control solution, Bottleneck Bandwidth, and Round trip propagation time congestion control algorithm (BBR) was proposed in 2016 by Cardwell N. et al. in [32]. Authors have used the term “congestion-based” congestion control, meaning the aim of the algorithm is to operate near Kleinrock’s optimal point

(see Fig. 2) reaching high channel utilization yet keeping low queuing delay across the network. It uses periodic delay and bandwidth probing to evaluate the current buffering situation on a channel. The typical BBR flow behavior is presented in Fig. 5. In the contrast to loss-based solutions, this approach allows keeping a significantly lower bottleneck queue load level. This bandwidth probing mechanics leads to periodical short-time decreasing of the throughput. The network delay probing leads to periodical short-time increasing of the queue load level. The use case scenario is identical to the one described in Sect. 4.1.

As is shown in Fig. 6 TCP BBR demonstrates higher performance in most cases in comparison with the TCP CUBIC algorithm. The difference between these approaches is the increased throughput in the cases with packet loss probability (see Fig. 4 to compare). The testbed and emulation scenario are identical to those described in Sect. 4.1. Yeong-Jun Song et al. published research in 2020 in [33] demonstrating performance issues of BBR in shallow buffers and during a coexistence with other data streams in the same link. However, BBR seems to be an effective replacement of loss-based algorithms increasing the channel utilization in the cases with packet losses and decreasing the negative impact on the network.

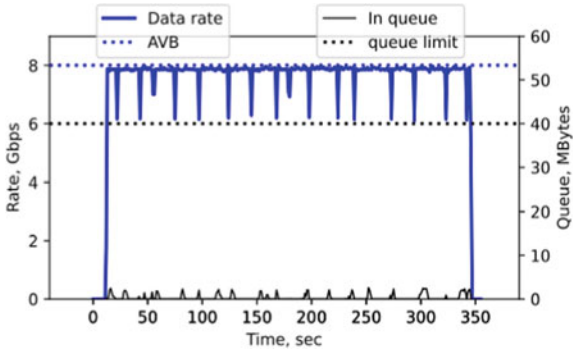



**Fig. 5** TCP BBR flow example

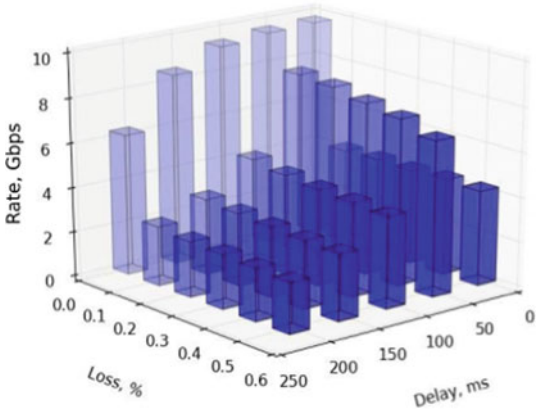



**Fig. 6** TCP BBR throughput

Addressing the same data transmission issues another solution has been proposed in [34]—the Bottleneck Queue Level congestion control (BQL). On the one hand, this approach is close to FAST TCP in the way of controlling the bottleneck queue. It evaluates the amount of data in the queue using the network delay metrics during data transmission and tends to keep this congestion at a constant level using PID control. Keeping the queue from being idle, this algorithm allows fully utilizing the bandwidth during the data transmission. Additionally, BQL uses one-way delay metrics instead of RTT what allows to avoid the negative impact of the congestion link in the acknowledges flow direction. On the other hand, this solution is close to the TCP BBR in the tendency to operate near the BDP of the link. In [10] was shown that the periodical bandwidth probing applied in TCP BBR can lead to significant performance issues in high bandwidth-delay product networks compared to smooth PID control applied in BQL. In Fig. 7 the difference between mentioned approaches is shown.

The use case can be described with the next emulation parameters: network delay equal 100 ms, available bandwidth equal 1 Gbps, and no packet losses. This experiment demonstrates the difference between TCP BBR and RMDT BQL approaches. While RMDT BQL tends to keep constantly full bottleneck bandwidth utilization, TCP BBR decreases its send rate periodically during the probing. While RMDT BQL keeps node queue load levels low and constant, TCP BBR periodically occupies the queue with short splashes. These differences in the behavior result in differences in the mean performance of the approaches. Figure 8 demonstrates the performance drop of the TCP BBR in comparison to the RMDT BQL in the same use cases.

Long pipes were emulated by Netropy with the same for each test available bandwidth equal to 1 Gbps. This parameter has been chosen because of limitations in the network stack, which allows comparing algorithms in more equal conditions. Results

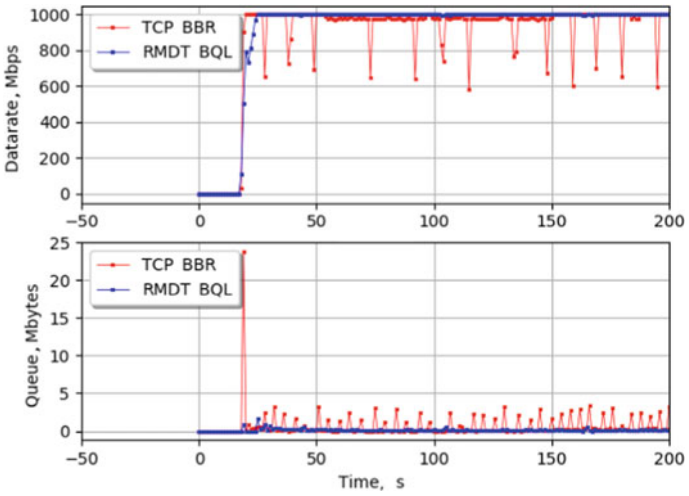



**Fig. 7** BQL/BBR 100 ms use case

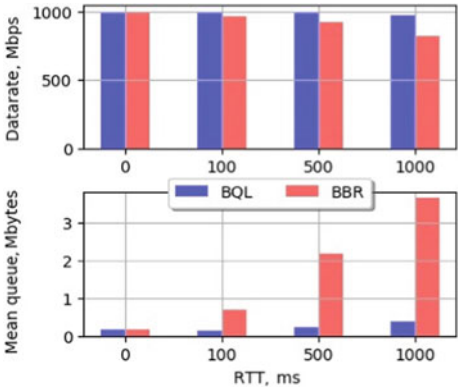



**Fig. 8** BQL/BBR summary

of similar series of tests but with higher available bandwidth as part of wide performance evaluation are presented below in Sect. 5.2. The drop of the mean throughput of TCP BBR controlled traffic can reach up to 15% compared to the RMDT BQL and the mean bottleneck buffer load level can be up to 10 times higher than using RMDT BQL in the worst use case provided. Such low requirements of the queue allow RMDT BQL to work in shallow buffers more effectively while TCP BBR can provoke packet losses in such an environment [33]. However, such a delay-based approach of congestion control as BQL can provide less performance in wireless networks. The reason is in dynamically changing parameters of the wireless link, which leads to ineffective link parameters evaluation as long as the metrics are going to be outdated rapidly. Despite the existing issues of congestion-based algorithms, it is a promising direction of congestion control researches.

## *4.3 Multipoint Data Delivery*

Group communications in TCP/IP networks are mostly applied with network-assisted multicast topology e.g. with IGMP and PIM protocols. Besides an existing and configured multicast architecture, reliable multicast application has to imply ARQ or/and FEK functions for every multicast client. In [35], Bakharev A. et al. have described problems of the multipoint data transmission protocols: congestion control, which is observed in this section, and, besides this, loss recovery and I/O buffer management that will be discussed below in Sect. 5. Described before point-to-point congestion control algorithm is can be used in a point-to-multipoint case with additional management over the receiver group management: an algorithm to evaluate the receiver with the lowest available bandwidth is required. This solution is reliable and efficient on small multicast groups but things are changing in the case when multicast groups consist of hundreds of receivers or more. When one particular source is suffering from a multicast feedback channel the overall data transmission performance is degrading significantly. It is possible to increase the performance of

data delivery by engaging network devices into the protocol base functionality. This allows to gather different metrics that correspond to a certain region of the distribution tree, increase retransmission performance, decrease the feedback channel, etc. The big drawback of such an approach is that such router-supported protocols require changes and additional configuration on the network nodes, which is frequently impossible. Moreover, multicast traffic can be blocked by the network provider. A possible solution for this issue is application-layer multicast (ALM). An application-layer multicast is a group of multipoint data transmission algorithms that offers multicast functionality without a multicast-enabled architecture using end-hosts for data replication and routing. In Fig. 9 the basic concept and architecture of ALM are shown.

A detailed survey of several application-layer multicast protocols was done in Computing Laboratory, the University of Kent [36]. In this work, existing algorithms were analyzed in terms of tree cost and delay optimization. Also, a classification by distribution tree principles for ALM is provided there. Later, Mojtaba H. in [37] did a detailed literature overview of different application-layer multicast approaches and their classification in terms of group configuration. Despite the advantages of the ALM, such an approach typically has several issues. Because of host-based retransmission of the data, ALM suffers from increased data delivery delays in comparing to the network-assisted multicast. This is related not only to increased path length but

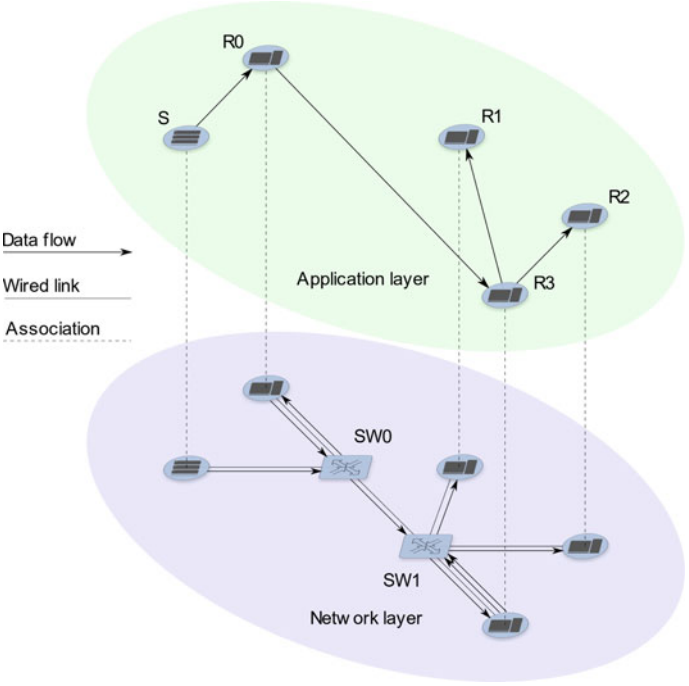



**Fig. 9** Application layer multicast topology

also to the necessity of buffering process on the retransmission nodes. Another significant problem is the dependency on the performance of every retransmission from the least effective node in the chain. It is more difficult to keep high data transmission performance of the ALM than network-assisted multicast because there are more disturbing factors in ALM, like hosts CPU, hosts I/O management, and number of links intended in the transmission. A UDP-based solution RMDT offers an application layer multicast functionality. Karpov K. et al. in [38] have proposed an adaptation of the Minimum Spanning Tree algorithm for the distribution tree construction using delay metrics as the main metric on the base of RMDT. This work is shown that such an approach can provide significantly higher data transmission rates than a comparable TCP-based multipoint data transmission realization. Additional work on this algorithm was published later in [39] describing available bandwidth evaluation as a second metric for distribution tree constructing for the purpose of increasing the data transmission performance in the presence of cross traffic.

## 5 Channel Utilization

Despite the number of modern congestion control and data transmission solutions proposed in the last years, the available bandwidth of the existing links is can be still not fully utilized. In addition to the issues of the transport protocols, limitations of modern networking stacks in operating systems result in bottlenecks on the sender or receiver side leading to low throughput. Standard Linux I/O calls are CPU-expensive what leads to a limited maximum packet per second rate for the concrete hardware. Also, Linux network stack is configured by default for relatively low-speed networks and it may be a problem to produce a high data rate to utilize the available bandwidth of some high-speed links with default network stack configuration. To overcome the limitation of single-threaded data transmission, several data streams over one link can be launched. This allows to engage more CPU threads into single data transmission and increase the possible amount of data in flight by separating it between several TCP buffers. This approach is proposed by Allcock W. et al. in a software solution called GridFTP [40] which is used in many projects that rely on high-speed TCP data transmissions over HBDP networks. Rashti M. J. et al. have observed in [7] how use of UDP instead of TCP in a GRID infrastructure can offload the NICs and decrease CPU usage and increase utilization of the channel. The significant drawback of such a solution is congestion control and resource sharing. Several data streams in the same link will challenge the resource sharing ability and robustness of all other connections in this link. Any other TCP connection entering this network will share network resources with many data streams of single data transmission. Such an approach makes this use case unfair and unacceptable for some scenarios. Modern data transmission protocols like RMDT include advanced mechanics and software solutions to increase the throughput controlled by one algorithm what makes this scenario more acceptable for wide area networks.

### 5.1 Packet Processing

Significant performance boosts have been achieved in RMDT with access-optimized memory buffer and distributing the process of packet allocation and parallel packet delivery to the kernel in parallel threads for a singular data transmission stream. The main idea of access-optimized buffer is to separate allocated memory into different segments with the predefined size equal to data packet size. These memory segments can be accessed directly by using the related sequence number of the data packet. This approach significantly decreases the time needed to access the data as long as no search is needed. However, this solution leads to not optimal memory usage, keeping unused data segments during data transmission. Syzov et al. in [11] have observed the scalability of the multithreading send routine and pointed that it is scaling well until the number of threads will reach the number of cores (or Hyper-threads) then, the overhead of context-switches decreases performance rapidly. Figure 10 demonstrates a significant sender performance boost resulting from implementing both solutions. The first two use cases here are related to 10 Gbps link and "standard" MTU equal 1500 bytes. Next two cases were tested in 40 Gbps link with "jumbo" MTU equal 9000 bytes. The most significant performance benefit is observed in the point-to-point case with enabled multithreading send routine. The send rate is nearly triple higher with the same hardware. Access optimized buffer enhancement plays an especially important role when protocol handles a high number of packets per second and it is more visible in 40 Gbps cases.

However, this approach is only part of the solution, since the receive side is still suffering from performance issues. To increase the maximum performance achievable on the single network socket, the packet receive routine has to be simplified. In general, it means decreasing the number of copy procedures needed to process each data packet and passing it directly to the users-pace. With the libraries like Data Plane Development Kit (DPDK) it is possible to significantly increase the packet per second receive rates. It allows to route data packets through shortest way in the host machine, decreasing the CPU cost on each receive call. D. Syzov et al. have shown achieved performance boost by the implementation of RMDT send/receive routines

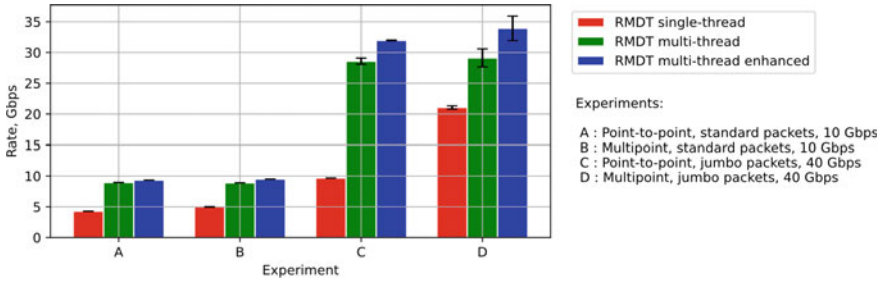



**Fig. 10** RMDT rate enhancing with multithreading and access optimized buffer

over DPDK library in [41]. Another approach is use of several sockets simultaneously within a multithreading send and receives operations. The approach differs from one used in TCP Grid, as long as the whole data stream is controlled by a single traffic control algorithm and doesn't break the coexistence in the channel. Mentioned solutions are described in more detail by Karpov K. et al. in [12].

### *5.2 TCP/IP Stack Tuning*

TCP performance in high bandwidth-delay product networks can be significantly increased by proper network stack parameters tuning. By default, Linux OS limits such parameters as socket buffer size and MTU size. Hanford N. and Tierney B. have demonstrated in [42] that TCP can reach high rates even in lossy high bandwidth-delay product channels with certain host configurations (e.g. nearly 9 Gbps in a channel with 100 ms delay and 1% packet losses depending on the hardware). In other work, Hock M. et al. demonstrated in [43] channel utilization of tuned TCP/IP stack different TCP congestion control in 100 Gbps network. They were able to reach 100 Gbps channel utilization in the use case with 20 ms network delay with 4 parallel TCP CUBIC streams.

In this work, a tuned TCP/IP stack was tested in a 10 Gbps link under different network conditions in order to examine its possibilities to utilize the lossy high bandwidth-delay product link with given hardware. The testbed is identical to the one used in 4.1. The host system was tuned up to satisfy the requirements of the use cases. First, the jumbo frames have been enabled and txqlen on interfaces have been extended. Second, a cpufreq profile was set to performance mode making it work on a higher frequency to provide the best performance. A list of other important settings that have been applied is presented in Table 1. The performance comparison of TCP BBR with tuned Linux network stack and RMDT BQL in the same network conditions can be observed in Fig. 11.

These tests include ranges of emulated network delay from LAN (tens of μs) to 500 ms; flat distributed probability of the packet loss from 0 to 1%; and the maximum available bandwidth equal 10 Gbps. The result of TCP BBR testing on tuned TCP/IP stack made by ESnet was reproduced - data stream reached nearly the same rate in the use case of 100 ms delay and 1% packet loss. However other PLR/RTT cases were resulted in lower performance down to nearly 3 Gbps in the worst case. Also, the significant performance boost related to TCP/IP stack tuning can be observed comparing Fig. 6 above in Sect. 4.2 with one presented in this section. RMDT, a UDP-based solution, in the contrast to TCP demonstrates high and stable channel utilization during the tests.

**Table 1** Linux TCP/IP Stack Tuning list for sysctl

| Parameter | Default | Tuned |
|---|---|---|
| net.ipv4.tcp_window_scaling | 1 | 1 |
| net.ipv4.tcp_timestamps | 1 | 1 |
| net.ipv4.tcp_sack | 1 | 1 |
| net.core.rmem_max | 212 992 | 1 073 741 824 |
| net.core.wmem_max | 212 992 | 1 073 741 824 |
| net.ipv4.tcp_rmem | 4 096 | 4 096 |
| | 87 380 | 87 380 |
| | 6 291 456 | 1 073 741 824 |
| net.ipv4.tcp_wmem | 4096 | 4096 |
| | 16,384 | 87,380 |
| | 4 194 304 | 1 073 741 824 |
| net.ipv4.tcp_mem | 140,964 | 1 073 741 824 |
| | 187,954 | 1 073 741 824 |
| | 281,928 | 1 073 741 824 |
| net.ipv4.udp_mem | 181,650 | 1 073 741 824 |
| | 242,203 | 1 073 741 824 |
| | 363,300 | 1 073 741 824 |
| net.core.netdev_max_backlog | 1 000 | 250 000 |
| net.ipv4.tcp_mtu_probing | 0 | 1 |
| net.ipv4.tcp_no_metrics_save | 0 | 1 |
| net.core.default_qdisc | fq_codel | fq |
| net.ipv4.tcp_congestion_control | cubic | bbr |
| MTU | 1 500 | 9 000 |
| txqueuelen | 1000 | 10 000 |
| CPU freq scaling | Balanced | Performance |

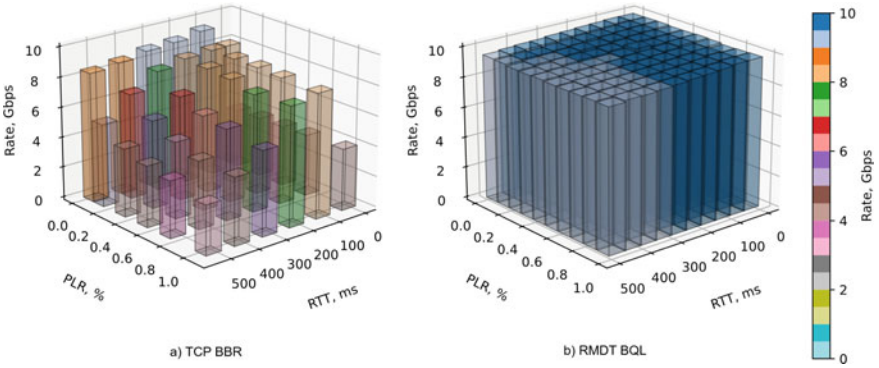



**Fig. 11** TCP BBR vs RMDT BQL in 10 Gbps throughput comparison

## 6 Conclusion

In this paper, several data transmission issues in high bandwidth-delay product networks have been discussed. A significant part of this work is devoted to network congestion and methods to avoid its negative impact. This is due to the fact that inefficient algorithms today decrease network performance globally. Most popular congestion control solutions nowadays tend to utilize the available network resources not optimally and provoke additional network delays and packet losses. WAN is not transparent for multicast traffic. Multicast infrastructure implies the control of the network nodes and complex network configurations. Moreover, multicast traffic is frequently blocked. This leads to difficulties and limitations for the usage of benefits of the multicast concept. The possible solution is application-layer multicast, which relies on controlled end hosts but not networks nodes. The second significant part of the work was devoted to the problem of underutilization of the existing highspeed channels. As long as the TCP stack is developed to work in the general case it can demonstrate low performance in high bandwidth-delay product networks. Besides this, the UDP-based solutions propose relatively high performance and resource-sharing features in contrast to TCP. Such solutions are not limited by one implementation of the data transmission protocol. It can be implemented with advanced algorithms and provide additional functionality decreasing the mentioned issues and allowing new ideas to come.

## References

1. Cisco: VNI Complete Forecast Highlights (2017–2022). White Paper (2018). https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html#~insights. Accessed 14 May 2021
2. He E, Leigh J, Yu O, Defanti TA (2002) Reliable Blast UDP: predictable high performance bulk data transfer. In: Proceedings. IEEE International Conference on Cluster Computing, pp 317–324
3. Meiss M, Tsunami R (2004) A high-speed rate-controlled protocol for file transfer. Indiana University
4. Gu Y, Grossman R (2007) Abstract UDT: UDP-based data transfer for high-speed wide area networks. Comput Netw 51:1777–1799
5. Ren Y, Haina T, Jun L (2009) Hualin Q (2009) Performance comparison of UDP-based protocols over fast long distance network. Inf Technol J 8(4):600–604
6. Se-Young Y, Brownlee N, Mahanti A (2013) Comparative performance analysis of high-speed transfer protocols for big data, pp 292–295, November 2013
7. Rashti MJ, Sabin G, Kettimuthu R (2016) Long-haul secure data transfer using hardware-assisted GridFTP. Futur Gener Comput Syst 56:265–276
8. Legrand I et al (2009) MonALISA: an agent based, dynamic service system to monitor, control and optimize distributed systems. Comput Phys Commun 180(12):2472–2498
9. Nadig D, Jung E-S, Kettimuthu R, Fosterz I, Nageswara Rao SV, Ramamurthy B (2018) Comparative performance evaluation of high-performance data transfer tools. In: 2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pp 1–6

10. Mareev N, Kachan D, Karpov K, Syzov D, Siemens E (2019) Efficiency of BQL congestion control under high bandwidth-delay product network conditions. In: 7nd International Conference on Applied Innovations in IT (ICAIIT 2019), p 5, Koethen, Germany
11. Syzov D, Kachan D, Siemens E (2016) High-speed UDP data transmission with multithreading and automatic resource allocation. In: 4nd International Conference on Applied Innovations in IT (ICAIIT 2016), p 5, Koethen, Germany
12. Karpov K, Kirova V, Mareev N, Syzov D, Siemens E (2021) Data transmission performance enhancement in multi-gigabit wide area networks. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technology and Systems. MCT 2019. LNNS, vol 152, pp 161–190. Springer, Cham. https://doi.org/10.1007/978-3-030-58359-0_9
13. Kleinrock L (1979) Power and deterministic rules of thumb for probabilistic problems in computer communications. In: International Conference on Communications, pp 43.1.1–43.1.10, June 1979
14. Villamizar C, Song C (1994) High performance TCP in ANSNET. SIGCOMM Comput Commun Rev 24(5):45–60
15. Appenzeller G, Keslassy I, McKeown N (2004) Sizing router buffers. In: Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM 2004, pp 281–292, New York, NY, USA
16. Gettys J (2011) Bufferbloat: Dark Buffers in the Internet, White Paper, October 2011. https://queue.acm.org/detail.cfm?id=2071893. Accessed 14 May 2021
17. Baker F, Fairhurst G (2015) IETF recommendations regarding active queue management. Request for Comments RFC 7567, Internet Engineering Task Force, July 2015
18. Adams R (2013) Active queue management: a Survey. IEEE Commun Surv Tutor 15(3):1425–1476
19. Floyd S, Jacobson V (1993) Random early detection gateways for congestion avoidance. IEEE/ACM Trans Netw 1(4):397–413
20. Kuzmanovic A, Ramakrishnan KK, Mondal A, Floyd S (2009) Adding Explicit Congestion Notification (ECN) Capability to TCP's SYN/ACK Packets. Request for Comments RFC 5562, Internet Engineering Task Force, June 2009
21. Blanton E, Paxson V, Allman M (2009) TCP congestion control. Request for Comments RFC 5681, Internet Engineering Task Force, September 2009
22. Afanasyev A, Tilley N, Reiher P, Kleinrock L (2010) Host-to-Host congestion control for TCP. IEEE Commun Surv Tutor 12(3):304–342
23. Paxson V, Allman M, Stevens WR (1999) TCP Congestion Control. Request for Comments RFC 2581, Internet Engineering Task Force, April 1999
24. Floyd S (2003) HighSpeed TCP for large congestion windows. In: RFC 3649, December 2003
25. Xu L, Harfoush K, Rhee I (2004) Binary increase congestion control for fast, long distance networks. In: IEEE INFOCOM, pp 2514–2524
26. Ha S, Rhee I, Xu L (2008) CUBIC: a new TCP-friendly high-speed TCP variant. ACM SIGOPS Oper Syst Rev 42(5):64–74
27. Mishra A, Sun X, Jain A, Pande S, Joshi R, Leong B (2019) The great internet TCP congestion control census. In: Proceedings of the ACM on Measurement and Analysis of Computing Systems, vol 3, no 3, pp 1–24, December 2019
28. Sheshadri RK, Koutsonikolas D (2017) On packet loss rates in modern 802.11 networks. In: IEEE INFOCOM 2017–IEEE Conference on Computer Communications, pp 1–9, Atlanta, GA, USA, May 2017
29. Brakmo LS, Peterson LL (1995) TCP Vegas: end to end congestion avoidance on a global Internet. IEEE J Sel Areas Commun 13(8):1465–1480
30. Wei DX, Jin C, Low SH, Hegde S (2006) FAST TCP: motivation, architecture, algorithms, performance. IEEE/ACM Trans Netw 14(6):1246–1259
31. Jamali S, Talebi M, Fotohi R (2021) Congestion control in high-speed networks using the probabilistic estimation approach. Int J Commun Syst 34(7):e4766 (2021)
32. Cardwell N, Cheng Y, Gunn CS, Yeganeh SH, Jacobson V (2017) BBR: congestion-based congestion control. Commun ACM 60(2):58–66

33. Song Y-J, Kim G-H, Cho Y-Z (2020) BBR-CWS: improving the inter-protocol fairness of BBR. Electronics 9(5):862
34. Mareev N, Kachan D, Karpov K, Syzov D, Siemens E, Babich Y (2018) Efficiency of a PID-based congestion control for high-speed IP-networks. In: 6nd International Conference on Applied Innovations in IT (ICAIIT 2018), p 5, Koethen, Germany
35. Bakharev AV, Siemens E, Shuvalov VP (2014) Analysis of performance issues in point-to-multipoint data transport for big data. In: 2014 12th International Conference on Actual Problems of Electronics Instrument Engineering (APEIE), pp 431–441, Novosibirsk, Russia, October 2014
36. Tan S, Waters G, Crawford J (2003) A survey and performance evaluation of scalable tree-based application layer multicast protocols
37. Hosseini M, Ahmed DT, Shirmohammadi S, Georganas ND (2007) A survey of application-layer multicast protocols. IEEE Commun Surv Tutor 9(3):58–74
38. Karpov K et al (2019) Adopting minimum spanning tree algorithm for application-layer reliable mutlicast in global mutli-gigabit networks, March 2019
39. Karpov K, Iushchenko M, Mareev N, Syzov D, Siemens E, Shuvalov V (2020) Available bandwidth metrics for application-layer reliable multicast in global multi-gigabit networks, p 5
40. Allcock W, Bresnahan J, Kettimuthu R, Link M (2005) The Globus striped GridFTP framework and server. In: ACM/IEEE SC 2005 Conference (SC 2005), pp 54–54, Seattle, WA, USA
41. Syzov D, Kachan D, Karpov K, Mareev N, Siemens E (2019) Custom UDP-based transport protocol implementation over DPDK. In: Proceedings of the 7th International Conference on Applied Innovations in IT
42. Hanford N, Tierney B: Recent Linux TCP Updates, and how to tune your 100G host. https://fasterdata.es.net/assets/Papers-and-Publications/100G-Tuning-TechEx2016.tierney.pdf. Accessed 02 Jan 2021
43. Hock N, Veit M, Neumeister F, Bless R, Zitterbart M (2019) TCP at 100 Gbit/s–tuning, limitations, congestion control. In: 2019 IEEE 44th Conference on Local Computer Networks (LCN), pp 1–9, Osnabrueck, Germany, October 2019

# Research of the Service Structure Influence on the Sensitivity Indicators of the Queuing System Characteristics with Priorities




**Leonid Uryvsky and Kateryna Martynova**

**Abstract** The content of the article focuses on the study of the problem of access to system resources at the channel level of the OSI model. The problem of access is the objective lack of resources (time, frequency, energy, channels) for users of telecommunications services. Uneven distribution of resources creates a shortage of them among low-privileged users and a surplus for high-privileged subscribers. The problem is exacerbated by the fact that the constant development of widely available technologies and the circumstances in which society exists encourages users to increase their demands for quality of service, and as a result - a conflict of user access to telecommunications resources. The task of the study of telecommunications channels at the channel level in the framework of applied information theory is to select and further use adequate models to describe and quantify the transmission systems of information from the standpoint of access to telecommunications channels. By identifying the characteristics of the telecommunications system and the corresponding QoS, it is possible to formulate recommendations for the rational construction and organization of information transmission systems at the channel level with a known set of resources of telecommunications channels.

**Keywords** Queuing system with priority service · Service structure · Queuing system characteristics · Sensitivity indicators

## 1 Introduction

The limitation of telecommunications resources creates conflicts not only among users but also between telecommunications service providers and users. In the first

L. Uryvsky (✉) · K. Martynova (✉)
Information-Communication Technologies and Systems Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Industrialnyi Lane 2, Kyiv 03056, Ukraine
e-mail: leonid_uic@ukr.net

K. Martynova
e-mail: scorpio.for.teen@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_15

case, users are fighting for access to services, in the second—the provider has to meet the requirements of all users while ensuring a steady growing revenue for services. However, the provider ought to understand the impact of the technological components of queueing systems to make a correct business model. Therefore, there is a need to study the dependence of QoS on the number of service devices and queues.

The object of research - queuing systems with an arbitrary number of service devices (servers) and queues in combination with the use of priority servicing.

Let's define the goals of each participant in the queuing system.

The purpose of the user (jobs, requirements): not to receive a denial of service, to spend as little time as possible, standing in line.

The purpose of the service system (servers and channels): to be in a state of forced downtime for as little time as possible.

The purpose of the queuing system analysis: to reach a compromise between the requirements of "customers" and the capacity of the servicing system. To do this, the efficiency indicators of the queuing system (QS) are computed through the characteristics of its functioning.

The most significant characteristics of QoS functioning are productivity, time characteristics of requirements' servicing (average waiting time, average service time), and probabilistic indicators (probability of failure, probability of acceptance for service) [2].

Productivity is the average number of requirements serviced by the system per unit time at a given quality of service, which is characterized by time and probabilistic indicators.

The subject of research-the sensitivity of the queuing system QoS to service structural changes.

The structure of the service provides for a certain number of service devices, seats in the queue, the discipline of service.

Sensitivity reflects the ratio of changes in service quality (performance, probability of failure, etc.) to changes in service conditions and relevant parameters of the queuing system: $\lambda$—the intensity of incoming requests, $\mu$—the intensity of service requests, N—number of devices, m—number of flows, r—number of seats in the queue.

The article aims to describe a unified complex analytical model for priority queuing systems with an arbitrary number of service devices and queues and to formulate recommendations on the feasibility of using mechanisms to change the structure of service in queuing systems depending on the need to achieve QoS.

To achieve this goal the following tasks are solved:

- Selection of a set of queuing systems models using priority service disciplines as the most flexible in terms of impact on service quality indicators, and their study on universality and sensitivity. Thus, examining the QS models for sensitivity, a more extensive analysis of the QoS dynamics is obtainable.
- Creation of a universal complex model of a queuing system with an arbitrary number of devices and places for queuing. Versatility is ensured by changing

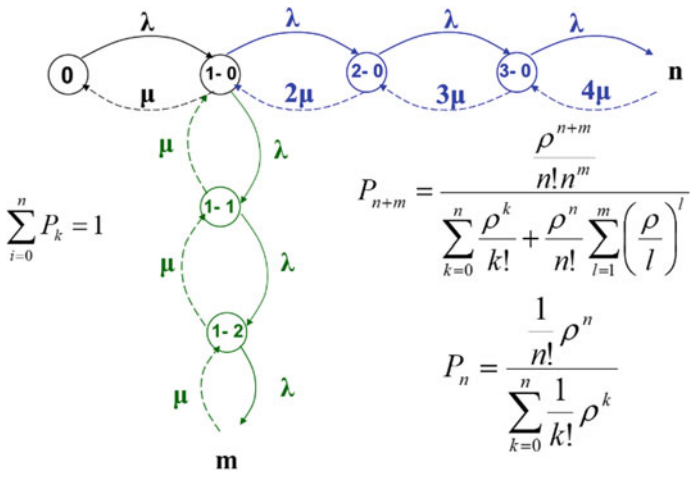



**Fig. 1** Queuing system with deterministic identification of states

the input parameters of the system ($\lambda_i$—the intensity of incoming applications, $\mu_i$—the intensity of application service), we obtain mathematical relations that correspond to one of the studied disciplines of service.

- Identification of intersection points in the graphic lines of the universality and sensitivity dynamics of QoS characteristics, based on which recommendations are formed for the application of priority service disciplines in the context of cloud computing.

In Fig. 1 provided a graph of QS states without priority maintenance with one input stream, the unlimited number of service devices, and seats in the queue.

The formula for the steady state of the system (absence of any application) is:

$$P_0 = \frac{1}{\sum_{k=0}^{n}\frac{\rho^k}{k!} + \frac{\rho^n}{n!}\sum_{l=1}^{m}\left(\frac{\rho}{l}\right)^l}, \quad (1)$$

where $k$—the index of the server, $n$—total number of servers in the system, $\rho$—the input load to the QS, Erlang, $l$—the index of the seat in queue, $m$—total number of seats in queue.

Remarkably, the formula for calculating the $P_0$ state of the graph is basic, because, substituting in the index of any denominator, you can get formulas for calculating the probability of any state of the QS.

The number of terms in the denominator of the expression for $P_0$ is equal to the number of states on the graph. In turn, each term is a path from $P_0$ to a specific state.

From Fig. 1 the following conclusion is formed: the numerator of the mathematical expression for calculating the probability of the state, which reflects the level of employment of the $n^{th}$ service device, is equal to $p_n = \frac{\rho^n}{n!}$ (with constant denominator), and the numerator of the expression for calculating the probability of the

application in the n + $m^{th}$ place for the queue, where n is the number of the service device, m is the number of places in the queue, equal to $p_{n+m} = \frac{\rho^{n+m}}{n!n^m}$ (with constant denominator) [4].

In a multi-streamed QS, there is also an access conflict when application services interfere over time. In a non-priority system, the advantage in service is always on the side of those applications that bring the maximum load on the system.

In reality, the most important jobs do not create extreme burdens. To ensure quality service of important messages, use priority service mechanisms - with absolute or relative priorities.

The essence of pre-emptive (absolute) discipline is that the requirement of higher priority interrupts the service of the requirement of lower priority [3].

Applications of the first type with an intensity of $\lambda_1$ and the parameter of exponential service $\mu_1$ are accepted as an upper-priority, and applications of the second type with parameters $\lambda_2$, $\mu_2$—lower-priority.

The following rules apply to jobs of the first type received by the device:

In case the device is idle, the application is accepted for service and receives it;

In the case when the device is busy servicing the application of lower priority, the interruption is carried out immediately, $\Delta t_0 = 0$.

If the device is busy servicing a higher priority application, the incoming job is lost.

The second type of jobs received by the occupied device is lost. The application received for a idle device is serviced to the end only if it is not interrupted by the the first type job, i.e. upper-priority.

The essence of relative priority (head of the line priority discipline) is that the requirement of higher priority is at the beginning of the queue without interruption [3].

Jobs of the first type with the intensity of receipt $\lambda_1$ and the parameter of exponential service $\mu_1$ are accepted as an upper-priority, and jobs of the second type with parameters $\lambda_2$, $\mu_2$—lower-priority. The following rules apply to the first type jobs received by the device:

- If the device is idle, the job is accepted for maintenance and serviced to the end;
- If the device is busy servicing another job of the second type, and the place in the queue is free, the job takes a place in the queue and is waiting to be serviced, after which it must be serviced;
- If the device is busy servicing the application of the first type of occupied seats in the queue, the incoming application is lost.

The second type of requirements received by the occupied device is lost. The requirement received for a free device is serviced to the end only if it is not interrupted by the first type requirement.

Thus, the use of priority service mechanisms gives any QS flexibility in achieving the required quality of service.

## 2 The Description of a Unified Complex Analytical Model for QS with an Arbitrary Number of Service Devices with Priority Servicing

The unified structure of QS with priority maintenance, which is characterized by increasing the number of service devices is represented in Fig. 2.

Theoretically, it is possible to increase the number of service devices indefinitely, but you should follow these rules:

1. At each iteration $i = 0..N$, where $N$ is the number of servers, the number of states at each new level increases by 1, from the previous level.
2. For each iteration $i = 0..N$, where $N$ is the number of servers, $\lambda = \text{const}$, but for each $P_N$ the intensity of service is defined as $N*\mu$.

The last level of the graph reflects the change in the state of the QS when jobs fight for the possibility of service through priorities because all available service devices are busy.

In this case, the states are divided into 2 types: the state of "donor" (application of lower priority) and the state of "recipient" (application of higher priority).

Figure 3 shows a similar generalized structure of the QS with priority service, one service device, which is characterized by increasing the number of seats in the queue.

Access conflict occurs when all the seats in the queue become occupied.

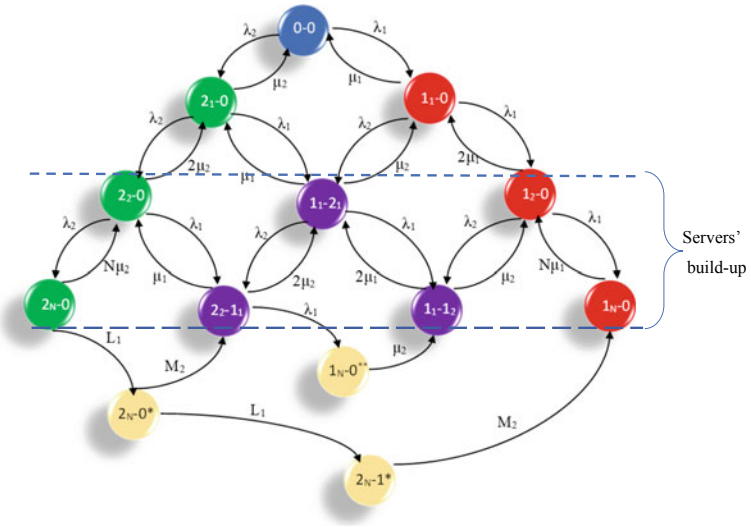



**Fig. 2** QS with priority service and servers' build-up

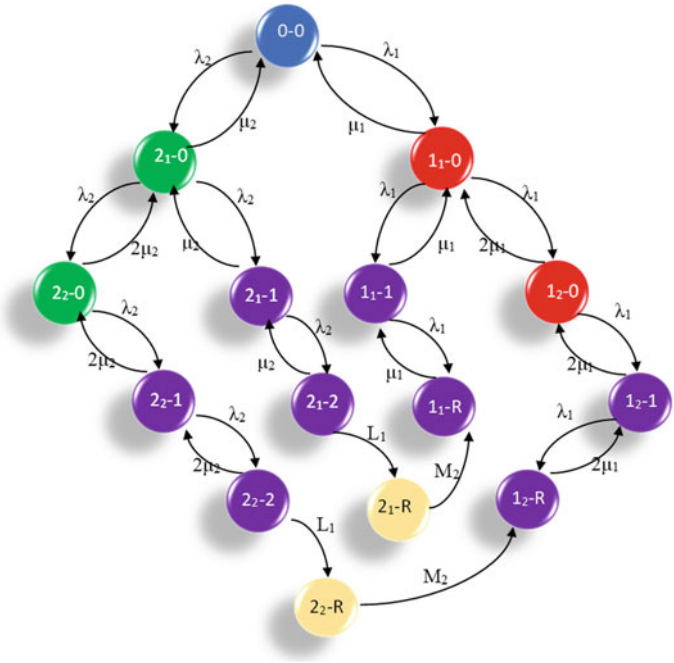



**Fig. 3** QS with priority service and queue build-up

Taking into account the method of determining the value of the probability of each state of QS, above, we build an analytical model for QS with an arbitrary number of service devices, an infinite queue, two streams, and priority service.

Formula (2) is a complex analytical model of QS with an arbitrary number of servers and queue and two jobs streams.

$$P_0 = \frac{1}{1+\sum_n\sum_r \frac{\rho_1^{n+r}}{n!n^r}+\sum_n\sum_r \frac{\rho_2^{n+r}}{n!n^r}+\sum_n\left(\frac{\rho_2^{n+R}}{n!n^r}\cdot\frac{1}{1+\frac{L_1}{n\mu_2}}\right)+\sum_n\left(\frac{\rho_2^{n+R}}{n!n^r}\cdot\frac{1}{1+\frac{L_1}{n\mu_2}}\cdot\frac{L_1}{M_2}\right)+\sum_n\left(\frac{\rho_1^{n+r}}{n!n^r}+\frac{\rho_2^{n+R}}{n!n^r}\cdot\frac{1}{1+\frac{L_1}{n\mu_2}}\cdot\frac{M_2}{n\mu_1}\right)}. \quad (2)$$

The denominator of the mathematical expression (2) is the sum of the factors $P_0$ to calculate the probability $p_{n,r}$ the state of the system, where $n = 1..N$ is the index of the server, $r = 0..R$ is the index of the seat in the queue.

The term 1 is a multiplier $P_0$ to calculate the probability of the state $P_0$.

The term $\sum_n \sum_r \frac{\rho_1^{n+r}}{n!n^r}$ is a factor of $P_0$ to calculate the total probability of engaging of all servers of the 1st stream.

The term $\sum_n \sum_r \frac{\rho_2^{n+r}}{n!n^r}$ is a factor of $P_0$ to calculate the total probability of employment of all servers of the 2nd stream.

The term $\sum_n \left(\frac{\rho_2^{n+r}}{n!n^r}\cdot\frac{1}{1+\frac{L_1}{n\mu_2}}\right)$ is a factor of $P_0$ to calculate the total probability of applications in the "donor" state.

The term $\sum_n \left( \frac{\rho_2^{n+r}}{n!n^r} \cdot \frac{1}{1+\frac{L_1}{n\mu_2}} \cdot \frac{L_1}{M_2} \right)$ is a multiplier $P_0$ to calculate the total probability of applications in the intermediate condition.

The term $\sum_n \left( \frac{\rho_1^{n+r}}{n!n^r} + \frac{\rho_2^{n+r}}{n!n^r} \cdot \frac{1}{1+\frac{L_1}{n\mu_2}} \cdot \frac{M_2}{n\mu_1} \right)$ is a multiplier $P_0$ to calculate the total probability of applications in the state of "recipient".

When $L_1 = \lambda_1$, $M_2 \to \infty$ we obtain a system with absolute priority. The system with relative priority meets the following conditions: $L_1 = \lambda_1$, $M_2 = \mu_2$. For the case when the number of servers increases, the following conditions must be observed: $n = 1..N\text{-}1$; for the case of queue: $r = 0..R\text{-}1$, where $N$ is the number of service devices, and $R$ is the number of seats in the queue [4].

Thus, the universal integrated analytical model of two-stream QoS with any number of service devices and queuing using priority service is a tool to determine all possible states of QoS, determine the characteristics of service, and a tool to change the service structure to achieve the required quality of service.

There are two basic scenarios for implementing the behavior of QS with priorities:

Both streams have the same input load, but the service strategies on the input parameters are diametrically opposed.

The priority flow is characterized by lower (less intense in the input flow and more intense in the output flow) quantitative values of the QS load.

## 3 The Research of Priority Queuing Systems' Sensitivity Characteristics to Service Structure Alterations

The "sensitivity" term is defined as "any change in a system function or a system characteristic caused by a change in one or more system parameters" [5].

To estimate the sensitivity of each indicator to changes in the value of the QS parameter, use the following method. Capture all parameters except one selected. Then remove the values of all indicators for several values of this selected parameter. Of course, the simulation procedure's repeated over and over again and averages the values for each value of the parameter to assess accuracy. But the result is a reliable statistical dependence of indicators on the parameter.

Then remove 12 more dependencies of indicators P on another parameter, say r - the number of places in the queue, fixing other parameters. And so on. A kind of matrix of dependences of indicators P on the parameter r is formed, which can be used for additional analysis of the prospects of movement (improvement of indicators) in one direction or another. The slope of the curves well shows the sensitivity, the effect of movement on a certain indicator. In mathematics, this matrix is called the Jacobian matrix J, in which the value of the slope of the curves is played by the values of the derivatives $P_i/R_j$, see Fig. 4.

It is worth noting that the curves were constructed based on the assumption that all but one of the parameters in the process of their removal were stable for the transparency of the research.

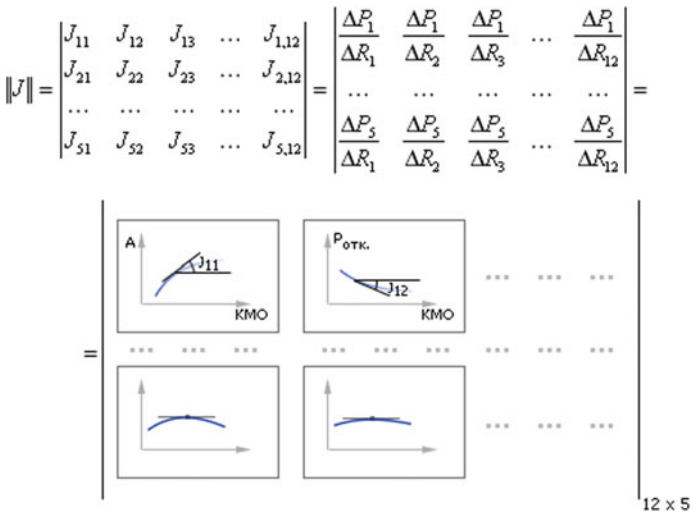



**Fig. 4** Jacobian matrix – the matrix of characteristics' sensitivity

Therefore, if a certain QS parameter decided to be changed on the basis of consideration of the removed curves, all curves for the new point, which will re-examine the question of which parameter should be changed to improve performance, should be removed again.

Performance is determined by Little's formula:

$$Y = \sum_{i=1}^{N} i * P_i, \tag{3}$$

where $N$—the number of servicing devices, $P_i$—the probability of system's being in $i$ state.

The sensitivity of performance $\Delta Y$ to changes of the devices number computes using the following formula:

$$\Delta Y_N = \frac{1 - \frac{\sum_{i=1}^{N} i * P_i}{\sum_{i=1}^{N-1} i * P_i}}{\frac{N}{N-1} - 1}. \tag{4}$$

The quantitative value of the average time spent by a request in the system computes using the Little's formula:

$$\tau_{sys} = \frac{\sum_{i=1}^{N} i * P_i}{\lambda}, \tag{5}$$

where $\lambda$—the intensity of requests income to system, $1/\tau_{in}$, $\tau_{in}$—the average time interval between incoming requests.

$$\Delta\tau_{sys_N} = \frac{1 - \frac{\sum_{i=1}^{N} i*P_i}{\sum_{i=1}^{N-1} i*P_i}}{\frac{N}{N-1} - 1}. \quad (6)$$

The quantitative value of the average number of requests in the system ($Z_{sys}$) computes using the Little's formula:

$$Z_{sys} = Y_{serv} + r_{queue}, \quad (7)$$

where $Y_{serv}$—serviced load, or performance, Erlang, $r_{queue}$—the average number of requests in queue.

In case of QS without a queue $r_{queue} = 0$.

Then sensitivity of average number of requests in the system $\Delta Z_{CHCT}$ computes using the following formula:

$$\Delta Z_{sys_N} = \frac{1 - \frac{\sum_{i=1}^{N} i*P_i}{\sum_{i=1}^{N-1} i*P_i}}{\frac{N}{N-1} - 1}. \quad (8)$$

Queue boosting. Performance for QS in turn is calculated by the formula:

$$Y = \sum_{i=1}^{N} \sum_{r=0}^{R} P_{i,r}, \quad (9)$$

where $R$ —number of requests in the queue.

The sensitivity of the performance indicator is computed by the expression:

$$\Delta Y_{N,R} = \frac{1 - \frac{\sum_{i=1}^{N} \sum_{r=0}^{R} P_{i,r}}{\sum_{i=1}^{N} \sum_{r=0}^{R-1} P_{i,r}}}{\frac{R}{R-1} - 1}. \quad (10)$$

The average number of requests in the system ($Z_{sys}$) is computed by the Little's formula [1]:

$$Z_{syst} = Y_{serv} + r_{queue}, \quad (11)$$

where $r_{queue}$ computed using expression (15) and $Y_{serv}$ computed using expression (9).

Thus, the sensitivity of the average number of requests in the system $\Delta Z_{sys}$ can be calculated as the sum of the sensitivity of the indicators of the service load and the average number of requests in the queue:

$$Z_{sys} = \Delta Y_{serv} + \Delta r_{queue}, \quad (12)$$

where $\Delta r_{queue}$ computed using expression (16) and $\Delta Y_{serv}$ computed using expression (10).

The quantitative value of the average time spent by a request in the system computes using the Little's formula:

$$\tau_{sys} = \frac{\sum_{i=1}^{N} \sum_{r=0}^{R-1} (i+r)*P_{i,r}}{\lambda}, \tag{13}$$

$$\Delta\tau_{sys_{N,R}} = \frac{1 - \frac{\sum_{i=1}^{N} \sum_{r=0}^{R-1} (i+r)*P_{i,r}}{\sum_{i=1}^{N} \sum_{r=0}^{R-2} (i+r)*P_{i,r}}}{\frac{R}{R-1} - 1}. \tag{14}$$

The average number of requests in the queue is computed by the Little's formula:

$$r_{queue} = \sum_{i=1}^{N} \sum_{r=1}^{R} (i+r)*P_{i,r}. \tag{15}$$

Thus the sensitivity of the average number of requests in the queue is computed by the Little's formula:

$$\Delta r_{queue} = \frac{1 - \frac{\sum_{i=1}^{N} \sum_{r=1}^{R} (i+r)*P_{i,r}}{\sum_{i=1}^{N} \sum_{r=1}^{R-1} (i+r)*P_{i,r}}}{\frac{R}{R-1} - 1}. \tag{16}$$

So, relations (3) … (16) determine the mathematical model, due to which it is possible to analyze the sensitivity of the QoS characteristics of QS with priorities for changes in the structure of service.

## 4 The Analysis of QoS Characteristics Sensitivity of QS with Priorities for Changes of Service Changes

An important role in identifying and optimizing key QS characteristics of telecommunications networks is taken by the sensitivity indicators of complex systems, which provide assessment and additional understanding of system behavior [5].

The object of the study is the QS with priority service. In particular, consider systems with relative and absolute priority services.

Figure 5 shows the dynamics of the sensitivity of the performance of the 1st and 2nd streams for priority QS depending on the change in the number of service devices.

The 1st stream of QS applications with relative priority was the most sensitive. The second stream of applications to the QS with absolute priority was the least sensitive. Accordingly, the greater the increase in the number of servers, the worse the performance, because the devices are idle [4].

Figure 6 shows the dynamics of the sensitivity of the average time of the application in the system from changes in the number of service devices.

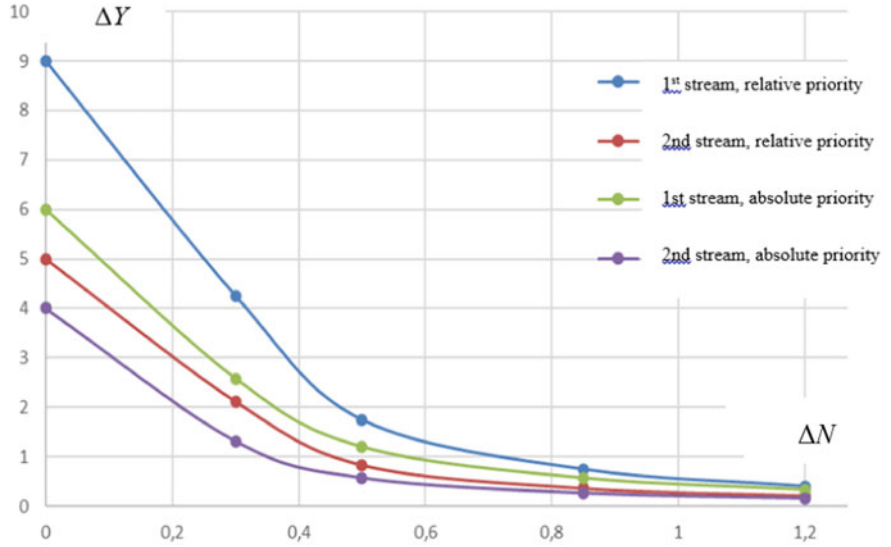



**Fig. 5** Dependence of the sensitivity of the performance depending on the change in the number of service devices

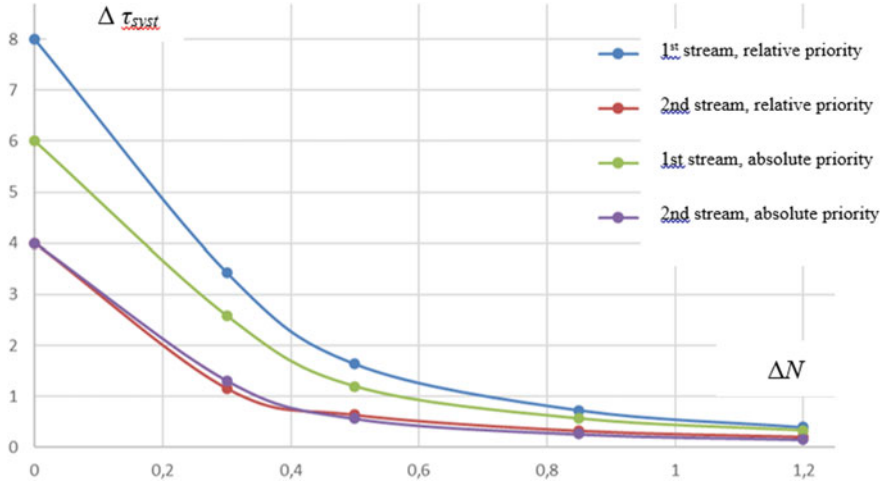



**Fig. 6** Dependence of the sensitivity of the average time of the request in the system depending on the change in the number of service devices

It is obvious that the advantage in the growth of $\Delta\tau_{syst}$ for applications of the first type. The queuing system with absolute priority has a 90% lower increase $\Delta\tau_{syst}$ due to the lack of seats for service waiting. On the one hand, the less time the application spends in the system, the faster it is serviced–so the user's need for fast service is met,

and the provider can service more requests. However, it should be noticed that QoS with the absolute priority of reducing the average time of the application in the system is achieved by rejecting applications of lower priority, i.e. for them, there is a high probability that $\tau_{sys} = 0$ s. So, on the other hand, users with lower priority requests receive a denial of service and, consequently, are dissatisfied with the services, and the provider loses a segment of customers, and hence revenue. Ultimately, even if only providers with priority requests will use the provider's services, there will again be a distinction between upper and lower priority users. Therefore, it is necessary to determine the optimal increase in devices, in which the system will be able to meet the requirements of users with all types of priority flows while maintaining the balance of time, i.e. the probability that the application $\Delta\tau_{sys} \neq 0$ was as high as possible.

Therefore, with a constant input load of the application and with the increasing growth of devices, the average length of stay of the application in the system $\tau_{syst}$ decreases. This fact is undoubtedly beneficial to the user services. Interestingly, that the curves $\tau_{syst}$ for job flows of the second type have an intersection point at $\Delta N = 0.4$. This circumstance is a prerequisite for further study of the average time of the application in the system ($\tau_{syst}$) in the context of situational management.

Obviously, service quality indicators are inextricably linked. Therefore, observing reduction the probability that $\tau_{syst} = 0$, we move on to the probabilistic characteristics of service quality, in particular to the probability of failure.

Consider the dynamics of the sensitivity of the probability of service depending on the build-up of service devices (see Fig. 7). The largest increase in the probability of service $P_{serv}$ is observed in the QS with relative priority (upper-priority stream). In second place was the priority flow of QS with absolute priority. In third place is the line of dynamics of QS sensitivity with the flow of lower priority using the relative discipline of service. And the least increase in probability $P_{serv}$ received a stream of jobs of the second type with the use of absolute service discipline.

Let's compare the obtained graph in numbers. It is obvious that the increase in the probability of servicing the first type of applications in the QS with relative priority is 120% higher than in the same flow of QS jobs, but with the use of absolute priority.

Consider the dynamics of the sensitivity of the average number of applications in the system depending on the change in the number of service devices (Fig. 8). Note that the indicators of the average number of applications in the system $Z_{syst}$ and the average residence time of the application in the system $\tau_{syst}$ are directly proportional values, so Figs. 5 and 8 similar.

The curves $\Delta Z_{syst}$ for the first type jobs do not intersect, in contrast to the curves for the second type jobs: there are two intersection points - at $\Delta N = 0.25$ and $\Delta N = 0.5$. That is, at $0.25 \leq \Delta N \leq 0.5$ QS with absolute priority is more sensitive in terms of $\Delta Z_{syst}$.

Summing up, we formulate the following conclusion: changing the number of servers is a tool for situational management of the service structure, moreover, the flows of higher priority applications are more sensitive to the servers build-up.

Consider the effect of queuing on the sensitivity of QoS characteristics. Figure 9 shows that lower priority application flows are more sensitive to queue build-up,

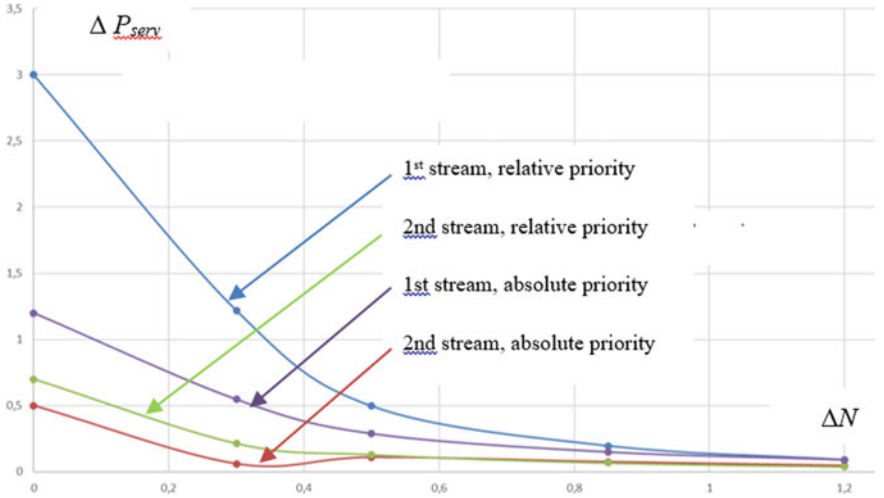



**Fig. 7** The dynamics of the probability of service sensitivity depending on the devices build-up

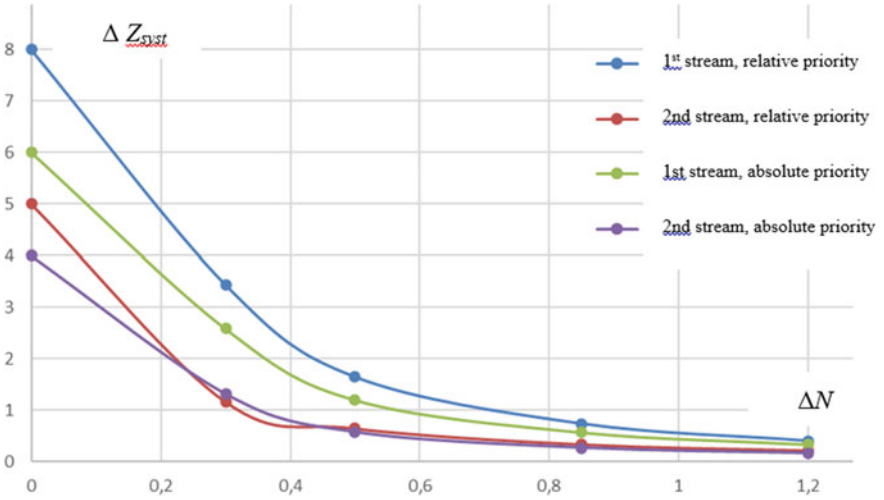



**Fig. 8** Dependence of the sensitivity of the average number of requests in the system depending on the change in the number of service device

especially QoS with relative priority. When increasing the number of seats for the queue to 30%, there is the largest difference in the values of the sensitivity indicator for the flows of applications of the first type –0.5.

Pay attention to Fig. 9, which shows the dynamics of the sensitivity of the productivity of QS with priority service disciplines. Obviously, the second type of jobs prevailed. Note that the requirements of the second type with relative priority has a

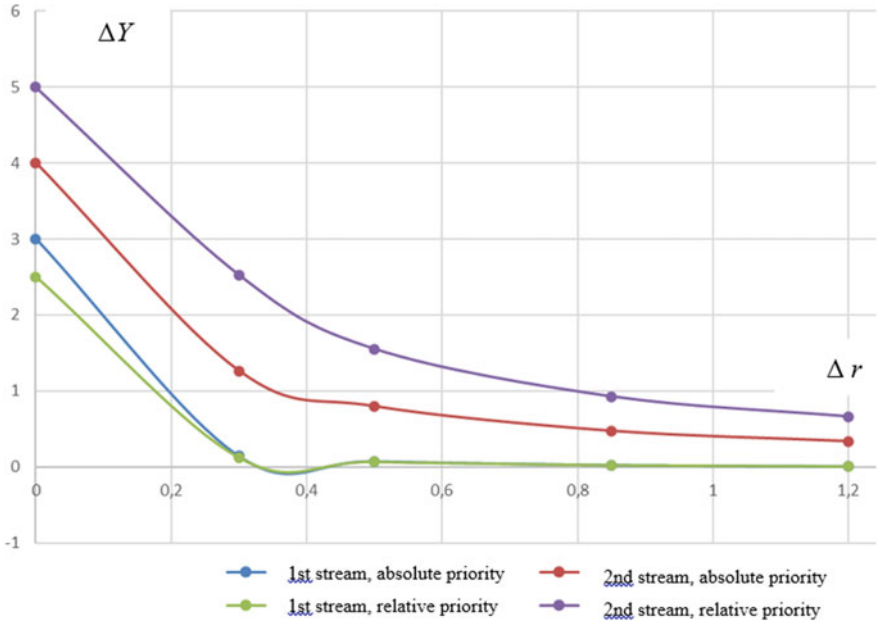



**Fig. 9** Dependence of productivity sensitivity depending on the change in the number of seats in the queue

150% higher increase in productivity than a similar requirement, but with the use of absolute priority. Comparing the increase in productivity for the first and second streams, the graph shows that the sensitivity of the second stream of applications to increase the queue is higher by 270%.

However, the opposite situation is with the upper-priority steams: the first stream of jobs in the QS with absolute priority productivity gains higher than in a similar flow with relative priority in the range from 50 to 0%. Moreover, the difference in productivity growth is leveled with increasing queue as soon as the number of waiting places increases by 30% ($\Delta r = 0.3$). Thus, it can be argued that queuing more than 30% is ineffective for managing the performance of upper-priority QSs.

Let's analyze the results from a practical point of view. The increased sensitivity of lower-priority queuing applications allows the provider to meet the needs of less privileged users without spending money on network equipment. However, the increase in productivity of lower-priority applications is obtained by reducing the increase in productivity of higher-priority applications, which will be a prerequisite for the conflict of users for access to services. Moreover, it has been determined that increasing queue space by more than 30% queue building becomes ineffective for managing the quality of service of priority flows.

The question also arises, how much time are heterogeneous users willing to spend waiting for service? How long can you increase the queue so that the waiting time does not exceed the time of failure, or the user does not leave the QS? In order to

answer these questions, it is necessary to analyze the sensitivity of time parameters of QoS service quality.

In Fig. 10 shows the dynamics of the sensitivity of the average time spent by the request in the system depending on the change in the number of seats in the queue.

Consider Fig. 10, which shows the dynamics of the sensitivity of the average residence time of the application in the system depending on the queue build-up. From the graphic lines, it is obvious that the priority was again given to applications with lower priority. Comparing the sensitivity of the average time of the application in the system by service disciplines, without a doubt, QoS with relative priority has a higher increase $\Delta\tau_{syst}$ due to the absence of service interruption of the second type of application when entering the flow of lower priority type of the first type. The difference in increment is in the range of 50 to 0% with increasing queue increment for all types of flows. The difference is leveled at the point $\Delta r = 0.3$ for the second type jobs and at the point $\Delta r = 0.5$ of the first type jobs.

In conclusion, increasing the queue by more than 30% for the second type jobs and increasing the queue by more than 50% for the first type jobs is ineffective for managing the sensitivity of the average time spent in the system ($\tau_{syst}$).

Let us formulate practical conclusions from the obtained results. In general, the increase in the average time of the application in the system is unfavorable for both users and providers, but it should be remembered that $\tau_{syst} = 0$–denial of service, which is also undesirable for both the user and the provider. Therefore, on the one hand, the advantage of the increase in $\tau_{syst}$ for lower priority applications means that the provider can serve more of the second type jobs, which will be waiting for service in the queue. However, as the queue increases, the sensitivity of the $\Delta\tau_{syst}$ index

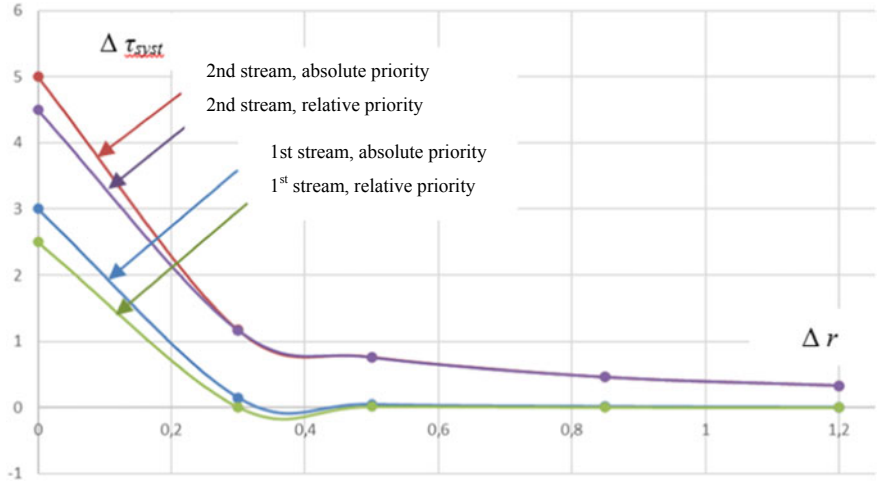



**Fig. 10** Dependence of the sensitivity of the average time of the request in the system depending on the change in the number of seats in the queue

decreases, i.e. this indicator becomes ineffective for service quality management. In particular, it was found that at the limit values of growth.

Let's look at Fig. 11, which shows the dynamics of the sensitivity of the probability of service depending on the increase in the queue. Again, there is a tendency for the predominance of the flow of applications of the second type of both disciplines of service. Moreover, the highest increase in the probability of service ($P_{serv}$) was the flow of lower priority applications in the QS with absolute service discipline. The maximum difference $\Delta P_{serv} = 175\%$, compared to a similar flow of QS with relative service discipline.

Analyzing the dynamics of the sensitivity of the probability of service ($P_{serv}$) for priority flows of applications, it can be argued that the QS with absolute priority has a higher increase (approximately 125%) in the probability of service than the QS with relative priority. However, with the increase in the queue within $0.3 \leq \Delta r \leq 0.5$, the increase $\Delta P_{serv}$ of the first stream of requests in the QS with relative priority is 15% higher than in the QS with absolute priority. Therefore, when increasing the queue by 30–50% of the initial number of seats in the queue, it is possible to control the increase in the probability of service between disciplines.

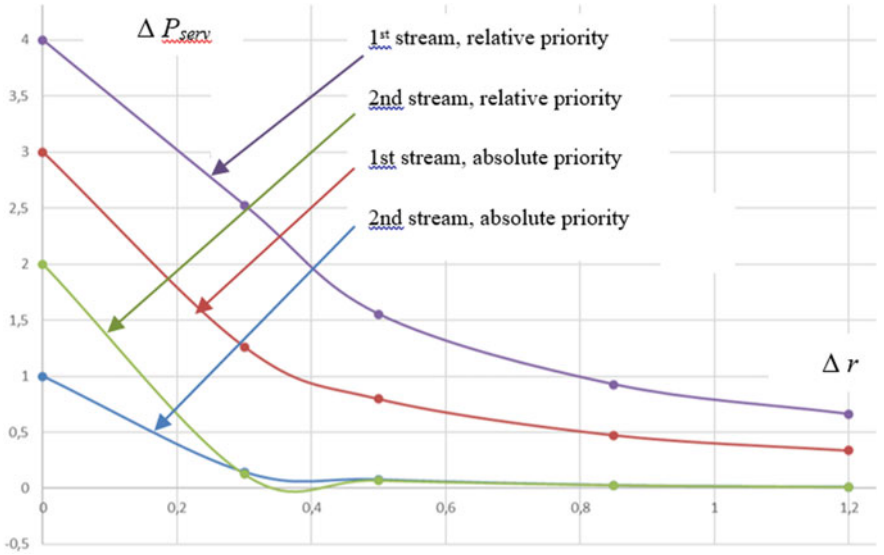



**Fig. 11** Dependence of the sensitivity of the probability of servicing depending on the change in the number of places for the queue

## 5 Analysis of Service Structure Management Scenarios in QS with Priorities

Calculated according to the above method, the table of choice of the structure of priority service, the operator of the service device has the opportunity to obtain an effective management procedure (see Table 1).

Based on the optimization criterion: maximizing the sensitivity of service quality indicators for each system service and minimizing the sensitivity of the probability of failure and the sensitivity of the time spent in the system - you can offer a number of recommendations to the service provider, namely:

1. The best tool to influence performance sensitivity for the first service is to increase the number of service devices and apply relative priority.
2. The best tool to influence the sensitivity of productivity for the second service is to increase the number of seats in the queue and apply relative priority.
3. The best tool to influence the sensitivity of the average time of the application in the system for the first service is to increase the number of service devices and the application of absolute priority.
4. The best tool to influence the sensitivity of the average time spent in the system for the second service is to increase the number of queues and apply absolute priority, provided that the number of waiting places does not exceed 30% of the initial queue length. In the study, the initial length of the queue is $R = 3$.

**Table 1** Analysis of the impact of QoS characteristics sensitivity with priorities on service structure management scenarios

| Characteristic | Stream | |
|---|---|---|
| | 1st stream ($\lambda_1$) | 2nd stream ($\lambda_2$) |
| $\Delta Y_1$ | Servers' build-up + relative priority | Queue build-up + relative priority (if $\Delta r \leq 30\%$) |
| $\Delta Y_2$ | Servers' build-up + relative priority | Queue build-up + relative priority |
| $\Delta \tau_{syst_1}$ | Servers' build-up + absolute priority | Queue build-up + relative priority (if $\Delta r \leq 50\%$) |
| $\Delta \tau_{syst_2}$ | Servers' build-up + absolute priority | Queue build-up + relative priority (if $\Delta r \leq 30\%$) |
| $\Delta P_{serv_1}$ | Servers' build-up + absolute priority | Servers' build-up + absolute priority (if $\Delta N \leq 50\%$) |
| $\Delta P_{serv_2}$ | Queue build-up + relative priority (if $\Delta r \leq 30\%$)\absolute priority (if $30\% \leq \Delta r \leq 50\%$) | Queue build-up + relative priority |
| $\Delta P_{refus_1}$ | Servers' build-up | Queue build-up |
| $\Delta P_{refus_2}$ | Queue build-up | Queue build-up |

5. The best tool to influence the sensitivity of the probability of servicing the application for the first service is to increase the number of service devices and the application of absolute priority.

To get a holistic picture of the quality of QS service using the priority disciplines of service, we will divide the concept of service quality, or QoS, into three dimensions: load, time, and probability measurements [6, 7]. There are two tools in the arsenal of influencing the outlined dimensions: queuing and instrument building.

Thus, the universal recommendation of QS with priority structure choice for telecommunication service provider are formulated based on complete analysis quantitative sensitivity QoS characteristics.

## 6 Conclusions

1. The object of research is the QS with priority service. In particular, consider systems with relative and absolute priority services.
2. An important role in determining and optimizing key performance indicators of telecommunications networks is taken by the sensitivity indicators of complex systems, which provide assessment and additional understanding of system behavior.
3. With a constant input load of the application and with increasing growth of servers, the average length of stay of the application in the system $\tau_{syst}$ decreases. This fact is undoubtedly beneficial to the user services. It should be noted that the curves $\tau_{syst}$ for application flows of the second type have an intersection point at $\Delta N = 0.4$. This circumstance is a prerequisite for further study of the average time of the application in the system ($\tau_{syst}$) in the context of situational management.
4. Changing the number of service devices is a tool for situational management, moreover, the flows of higher priority applications are more sensitive to the build-up of devices.
5. Increased sensitivity of lower priority requests to queuing allows the provider to meet the needs of less privileged users without spending money on the purchase of network equipment. However, the increase in productivity of lower-priority applications is obtained by reducing the increase in productivity of higher-priority applications, which will be a prerequisite for the conflict of users for access to services.
6. It was found that changing the number of places in the queue significantly increases the sensitivity of QoS indicators of lower priority applications.

## References


1. Kleinrock L (1975) Queueing Systems, Theory, vol 1, p 417. Wiley Interscience, New York. ISBN 978-0471491101. https://pdf.wecabrio.com/queueing-system-leonard-kleinrock.pdf
2. Uryvsky L, Martynova K (2019) Complex analytical model of priority requires service on cloud server. In: International Conference Radio Electronics & Info Communications (UkrMiCo). IEEE Xplore Digital Library, Scopus, pp 1–4. https://ieeexplore.ieee.org/document/9165323, https://doi.org/10.1109/UkrMiCo47782.2019.9165323
3. Uryvsky L, Gakhova A (2017) Scenario of the realization of situational priority for access systems. In: XI International Scientific Conference Modern Challenges in Telecommunications, Proceedings, pp 66–68. Institute of Telecommunication systems, Igor Sikorsky Kyiv Polytechnic Institute, Kyev
4. Martynova K, Uryvsky L (2021) The research of priority queuing systems' sensitivity characteristics to service structure alterations. In: Proceedings of the XV International Scientific and Technical Conference Modern Challenges in Telecommunications, pp 50–52. Institute of Telecommunication systems, Igor Sikorsky Kyiv Polytechnic Institute, Kyev
5. Toroshanko Y (2016) Management reliability of telecommunication network on the analysis of sensitivity of the complex systems, no 3, pp 31–36. Telecommunication and information technologies. http://stratum.ac.ru/education/textbooks/modelir/lection30.html
6. Uryvsky L (2009) Generalization of the Birth-Death process on systems with access conflict, no 1, vol 7, pp 88–102. Research and production collection Scientific notes US&RIC
7. Kleinrock L, Gail R (1996) Queueing Systems: Problems and Solutions, p 240. Wiley-Interscience. ISBN: 978-0-471-55568-1. https://www.wiley.com/en-au/Queueing+Systems%3A+Problems+and+Solutions-p-978047155568

# Improving the Accuracy of User Location in the Wi-Fi Network Using Complex Spline-Functions

**Irina Strelkovskaya, Irina Solovskaya, and Juliya Strelkovska**

**Abstract** The rapid development of various applications and services (Skyhook, Wi2geo, Google maps, etc.), which function on the basis of determining the current location of the user, both global GPS, and local LPS, today require the development of new improved methods. This applies, above all, to methods for determining the local location of users indoors under conditions of high concentration of users and the difficulties in radio signal distribution. Considered the use of local methods of determining the location based on access point AP equipment within the Wi-Fi/Indoor network infrastructure. A comparison of known methods for determining the local location of users, based on different principles, namely, the AOA method, based on triangulation by signal level, RSS method based on RSSI signal power level measurement and TOA method based on distance trilateration to AP. It is shown that to improve the accuracy of determining the location of the user is appropriate to use a combination of several methods, which will remove the disadvantages of one method, supplementing the advantages of another. The main methods of determining the location in the Wi-Fi/Indoor network, in particular the Fingerprinted method, are considered. A new method for determining the user's location based on the finite element method using complex planar splines is proposed. An example of finding the values of the user positioning error in the Wi-Fi/Indoor network is shown. The use of the proposed method will improve the accuracy of determining the coordinates of the location of the user based on the AP equipment in the Wi-Fi network and ensure the provision of LBS-based services and applications to indoor users under different conditions of their provision.

I. Strelkovskaya (✉) · I. Solovskaya (✉)
International Humanitarian University, Rishelievska Street 28, Odesa 65000, Ukraine
e-mail: i.strelkovskaya@mgu.edu.ua

I. Solovskaya
e-mail: i.solovskaya@mgu.edu.ua

J. Strelkovska (✉)
National University "Odesa Law Academy", Fontanska Road Street 23, Odesa 65009, Ukraine
e-mail: 4800632s@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_16

**Keywords** Location · Wi-Fi/Indoor · RSSI · Access point · Power · Fingerprinted method · Finite element method · Linear complex planar spline · Accuracy

## 1 Introduction

The current technological level of development of radio access networks today involves the use of services and applications by users that are based on the user's location LBS (Location-based Services), such as: local search, determination of nearby infrastructure points, geolocation of photos, videos, targeted advertising based on location. Most of these applications are implemented on smartphones, so the location algorithms used often have significant and severe limitations in terms of power, memory, security, and computing resources. Mobile applications use LPS (Local Positioning System) local positioning systems based on Wi-Fi (IEEE 802.11n/ac/ad), ZigBee or Bluetooth (IEEE 802.15x) technologies. Particular attention in such networks is given to security and location determination [1–10].

It's known [1–5] that a user's location can be defined in both global and local coordinates. The undisputed advantage of GPS global positioning is its full coverage and high positioning accuracy, but only in outdoor areas. On the contrary, a local positioning system (LPS) allows accurate positioning indoors. Current LBS-oriented services (Skyhook, Wi2geo, Google maps, etc.) for monitoring employees, vehicles, prisoners, equipment in offices, industrial premises, logistics complexes, medical organizations, underground facilities, car parks. Existing positioning technologies meet an extensive range of requirements in different situations. However, there are still situations where the required positioning accuracy cannot be achieved, such as indoors, car parks, warehouses and workspaces.

Considering the basic principles of determining the location in a Wi-Fi/Indoor network based on AP (Access Point). It is known [1–5] that Indoor positioning is based on the physical characteristics of the radio signal and takes into account the RSS (Received Signal Strength) power level of the received signal and the user's location. The classification of indoor positioning methods is shown in Fig. 1 [5].

Consider the main Wi-Fi/Indoor positioning methods [1–5]:

- AOA (Angle of Arrival) using the angle signal received relative to the AP access point based on the antenna pattern;
- RSS—for measuring the power level of the received signal;
- TOA (Time of Arrival) and TDoA (Time Difference of Arrival) methods—on trilateration, according to which the user's location is determined according to measurements of the propagation delay of the radio signal between the user's mobile device and the AP access point.

Wi-Fi/Indoor radio access network positioning systems (Fig. 2) are based on three basic principles of functioning, triangulation, trilateration and scence analysis. The ToA and TDoA methods based on radio propagation delay measurement use

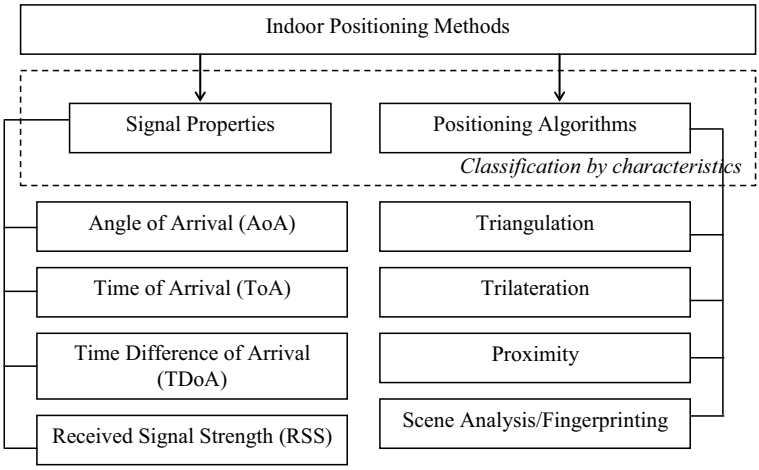



**Fig. 1** Classification of indoor positioning techniques [5]

trilateration, while the AoA method uses triangulation and the Fingerprinting method uses scence analysis.

The AoA method uses an angled signal received relative to an AP, based on the antenna pattern (Fig. 2a), to determine the user's location. APs equipped with a rotating antenna or phased array antenna are used for this purpose. In a Wi-Fi/Indoor network that uses the AoA method to determine a user's location, the accuracy of the location depends on the number of APs. The main advantage of the AoA method is the simplicity of the user location algorithm and the ability to operate on different physical principles. The disadvantages are the antenna design requirements and the low accuracy of user positioning.

The RSS (Received Signal Strength) method is based on the measurement of received signal strength and allows the location of a device to be determined based on the signal strength received from the AP (Fig. 2b). The main advantage of this method is the low power consumption of the mobile device when positioning and the low cost, but an important disadvantage is the low accuracy of positioning for significant distances. In fact, this method works well at short distances, but has significant errors when the range is increased due to the specifics of radio signal transmission.

The TOA (Time of Arrival) and TDoA (Time Difference of Arrival) methods are based on trilateration, whereby a user's location is determined according to radio signal transmission delay measurements between the user's mobile device and the AP (Fig. 2c). The TOA method is based on measuring the signal travel time from the device to the AP, where the distance to the target is calculated based on the difference between the time the signal is sent and the time it is received. However, this method requires strict time synchronization at the source and destination, which is difficult to obtain. The TDoA method measures the signal arrival time difference from a device to multiple access points. As with the previous method, strict time synchronization

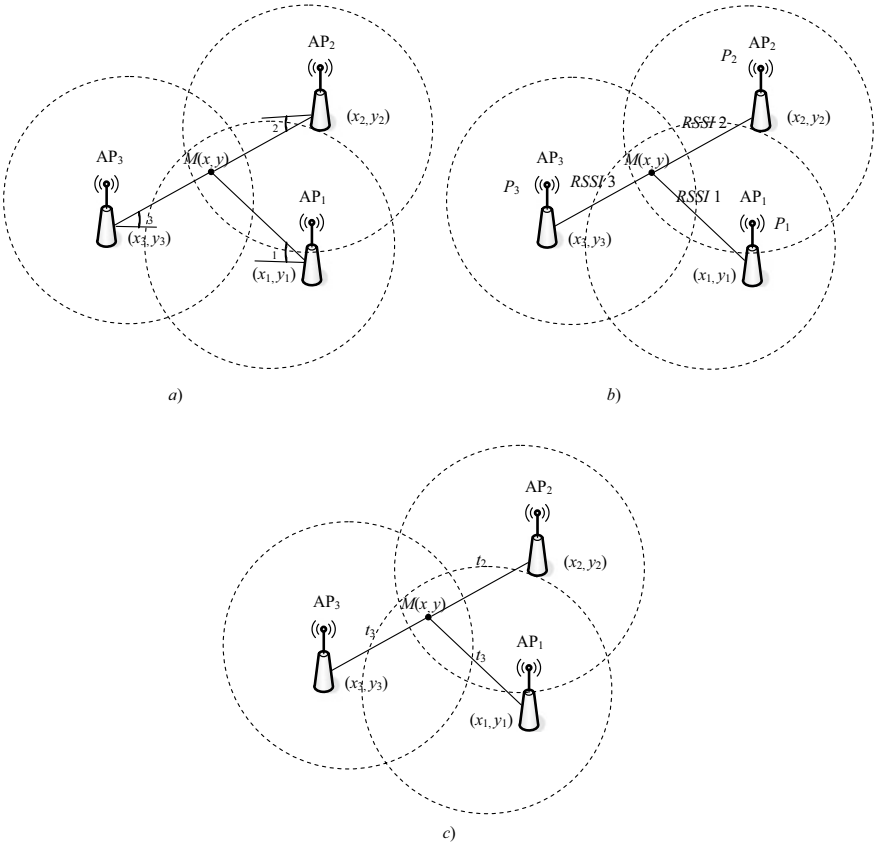



**Fig. 2** Methods for determining the location of the user on the Wi-Fi/Indoor network: **a** AoA method, **b** RSS method, **c** TOA and TDoA methods

is required, but only at the access points, there are no such requirements for user devices. An important advantage of these methods is the significant range.

The Fingerprinting method is based on determining the user's location based on the currently measured RSS signal strength values from all available APs with the values stored in a, pre-generated, database. The main advantage of the method is the accuracy of the user's location [4].

Along with the above methods, the Fingerprinting method is used, which makes it possible to assess the user's position by comparing the current measurements of the RSS power indicator of the received signal from the access points with the database of "radio fingerprints" of the signal power values corresponding to the values in the given coordinates [4].

Analysis of published works [1–5, 11–13] showed that methods based on time indicators and distance measurements in the "nearest" electromagnetic field NFER (Near-field Electromagnetic Ranging) are used to determine the user's location. Each of the above methods allows you to determine the location of the user with the required accuracy.

According to [11–13], the methods of a neural network work most accurately, but such a network must be trained every time the network changes. The overwhelming majority of works [1–4, 11–13] prefer the Fingerprinting method. The choice of this method is due to the fact that it is the easiest to implement, does not require additional equipment and the development of additional software, but allows achieving certain accuracy indicators. To improve the accuracy of determining the user's location, we will consider a modified method based on the Fingerprinting method and the finite element method using spline functions.

Earlier, the authors in [15–18] when solving problems of restoration and assessment of data states, signals and traffic, signal processing tasks and improving the quality characteristics QoS/QoE in the operation of information communication networks and predicting the characteristics of traffic of various applications, were used real spline functions (linear, quadratic, cubic, B-splines), which simplified the solution of many problems. However, there are a number of problems that cannot be solved using such splines, for example, the problem of determining the user's location in a Wi-Fi/Indoor network. In solving this problem, we will use complex planar splines [14, 19, 20].

The purpose of this work is to develop a method for determining the user's location in a Wi-Fi/Indoor network based on the Fingerprinting method using the finite element method based on complex planar splines, which will improve the accuracy of positioning.

## 2 Using the Fingerprinting Method on the Basis of Complex Planar Splines in Determining the User's Location in the Wi-Fi/Indoor Network

Consider a Wi-Fi/Indoor network (Fig. 3), which is a set of $AP_i$, access points, where $i$ is the number of $AP_i$ access points in the radio access network, and $i = \overline{1, m}$, $AP_i$ $(x,y)$ are the coordinates of the $i$ access point $AP_i$. [14].

Considering the finite element method, we use triangulation to determine the user's location in the Wi-Fi/Indoor network. According to triggered positioning, we will assume that at each point of the considered Wi-Fi/Indoor network, the user's device is within the range of at least three $AP_i$ access points.

Then the access points AP have coordinates: $AP_1$ $(x_1, y_1)$, $AP_2$ $(x_2, y_2)$, $AP_3$ $(x_3, y_3)$, and the coordinates of the user are $M$ $(x, y)$. Distances from access points $AP_i$, $i = \overline{1, 3}$ to user $M$ are equal $r_1$, $r_2$ and $r_3$, respectively.

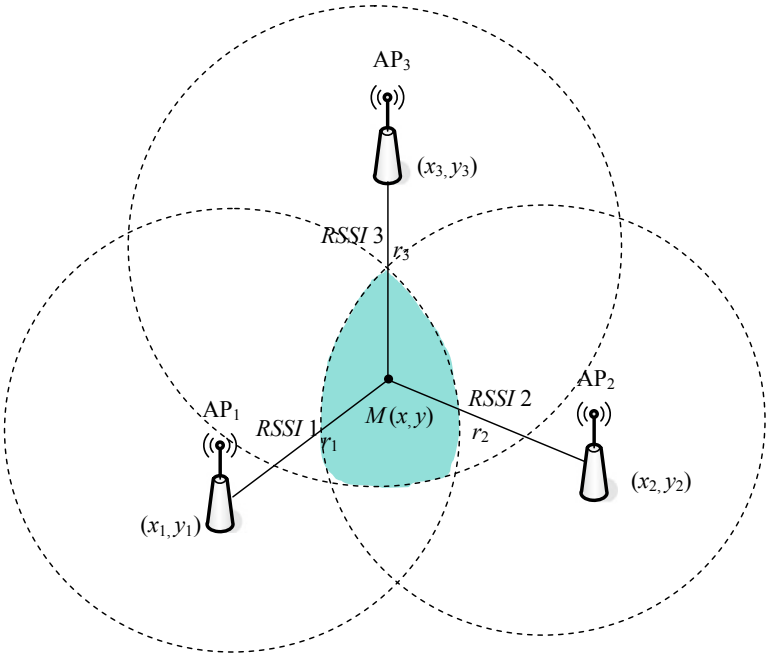



**Fig. 3** Wi-Fi/Indoor radio access network

To determine the distance between the user's device and the $AP_i$, $i = \overline{1,3}$ access points, the value of the received power level indicator RSSI is used according to the formula [4]:

$$P_{d_i} = P_0 - 10n \lg\left(\frac{d_i}{d_0}\right) \tag{1}$$

where $P_{d_i}$ is the value of the received signal strength RSSI of the corresponding access point $AP_i$, $i = \overline{1,3}$, $d_i$, $i = \overline{1,3}$ is the distance from the user device $M$ to the transmitter of the access point $AP_i$, $i = \overline{1,3}$, $d_0$ is the distance from the user device $M$ to the access point $AP_i$, $i = \overline{1,3}$ at which the signal strength was measured $P_0$, $P_0$ is the power signal, $n$ is the power loss factor of the signal during propagation in the medium.

Find the coordinates of user $M$ $(x, y)$ in the Wi-Fi/Indoor network using the Fingerprinted method based on triangulation using complex planar splines.

Consider the Fingerprinting method (Fig. 4) [14], in relation to determining the user's location in a Wi-Fi/Indoor network, comparing the currently measured values of the RSSI signal strengths from all available $AP_i$, $i = \overline{1,m}$ access points, with the values stored in a predefined network Wi-Fi/Indoor Data Base.

The operation of the Fingerprinting method consists of two stages. The first stage is the measurement of signal strengths at pre-planned known locations from all active

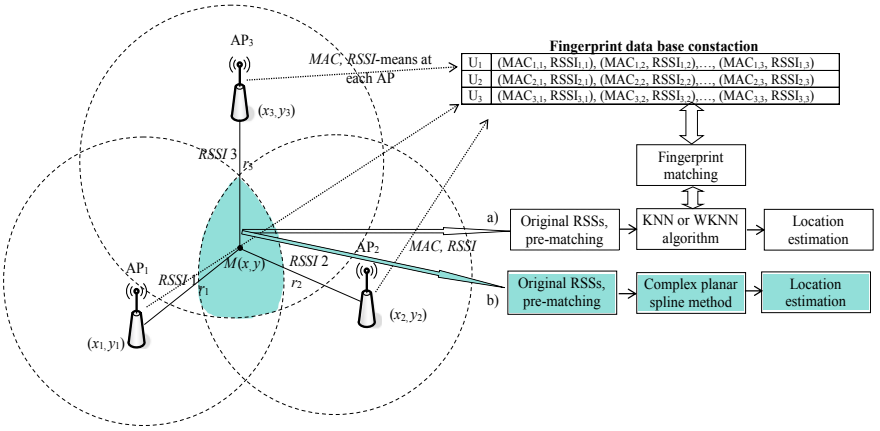



**Fig. 4** Fingerprinting method using the finite element method based on complex planar splines to determine the user's location in the Wi-Fi/Indoor network

APs. The collected information is stored in a database with reference to local room coordinates. The second step is location determination. At this stage, the signal strength measurements produced by the access point are compared with the information stored in the database by means of some algorithm for determining the coordinates of the user's location. For this, the method of finding the nearest neighbour NN (Nearest Neighbour) is used, which determines the user's coordinates as the arithmetic mean of the coordinates of the corresponding points, or the method of the nearest neighbour with the weighting coefficients WKNN (Weighted K-nearest) [4, 5].

Using the NN nearest neighbour method, one radio print nearest to the RSS measurement values is searched and it is decided that the current device coordinates match the coordinates of the corresponding room point stored in the DataBase [4, 5].

The K-nearest neighbour method (KNN) finds K radio prints for which the coordinates are calculated as the arithmetic average of the coordinates of the corresponding points. The K-nearest neighbour method differs from K-nearest neighbour method in that when calculating the coordinates, weighting coefficients are introduced, allowing to take into account the influence of the most distant or closest points on the result [4, 5].

The result of this step is the coordinates of the user's location. The obvious advantage of this algorithm is the location accuracy, which is achieved by a significant number of APs in the room. However, the Fingerprinting method has a number of disadvantages: the need for a large amount of time for database configuration and constant reconfiguration due to changes in the environment, the need for a significant number of AP equipment and high computational complexity.

For each user in the Wi-Fi/Indoor network, the Fingerprinting Data Base stores MAC-address data and the received power level values of the RSSI signal from the $AP_i$, $i = \overline{1, 3}$ access points, ($MAC_{1,1}$, $RSSI_{1,1}$), ($MAC_{1,2}$, $RSSI_{1,2}$), ($MAC_{1,3}$,

$RSSI_{1,3}$). Determining the user's location using the Fingerprinting method consists of two stages (Fig. 4) [14].

At the first stage, Fingerprint matching, RSSI measurements are performed from various $AP_i$, $i = \overline{1, m}$, access points, which are in the user's area $M$. At the second stage, Original RSSs, pre-matching according to the user's location and current RSSI values, the user's coordinates are determined (Fig. 4a).

According to [4, 5], the WKNN-weighted nearest neighbour method shows greater accuracy of position determination with a small number of points. If the access points are densely located, then greater accuracy is achieved using the NN method. However, both methods have some error in determining the user's location.

In this work, according to [14], along with determining the location of the user $M$ $(x, y)$ in the Wi-Fi/Indoor network, according to the above Fingerprinting method, we use a different approach based on the method of triangulation of the area under consideration using complex planar splines (Fig. 4b).

The Fingerprinting user positioning algorithm in a Wi-Fi/Indoor network is shown in Fig. 5. Let us consider a short description of the algorithm [4, 5]:

1. The initial step in the operation of the Location initiation algorithm is a Position request from the UE (User Equipment) to the nearest three $AP_i$, $i = \overline{1, 3}$.
2. Each $AP_i$, $i = \overline{1, 3}$, which has received a Position request performs signal detection and determines the Measure the RSS signal strength value and sends a Sending Fingerprint of its value for comparison to the previously created Fingerprint Data Base. The Data Base stores the measurement data of the RSSI power values from the $AP_i$, $i = \overline{1, 3}$, at various, predefined RP (Reference points), so-called "radio fingerprints".
3. The execution of the Position Algorithm after processing the information in the Data Base is based on the analysis of "radio fingerprints" and the received RSSI values.
4. To determine the coordinates of the user's UE location, the proposed finite element method based on the Complex planar spline is used.
5. As a result of calculations and construction of a linear complex planar spline, Position Result is obtained and the location of the user's UE in the Wi-Fi/Indoor network with coordinate's M $(x, y)$ is determined.
6. Finding the error in determining the coordinates of the location of the user UE Location Error is performed using Lemma 1.

To determine the coordinates of the user, consider linear complex planar splines [14]. Let a region $G$ of admissible values be given with nodes $z_1, z_2, \ldots, z_m$. The outer boundary of the Wi-Fi/Indoor network area is set. Thus, an area of the Wi-Fi/Indoor network covered by a network of access points $AP_i$, $i = \overline{1, m}$ and $s$-lines is created, divided into separate cells of arbitrary shape, including $m$-access points $AP_i$, $i = \overline{1, m}$ and $s$-lines. At each access point $z_j$ the signal strength $f(z_j)$ is known, according to which the user's location in the Wi-Fi/Indoor network is determined.

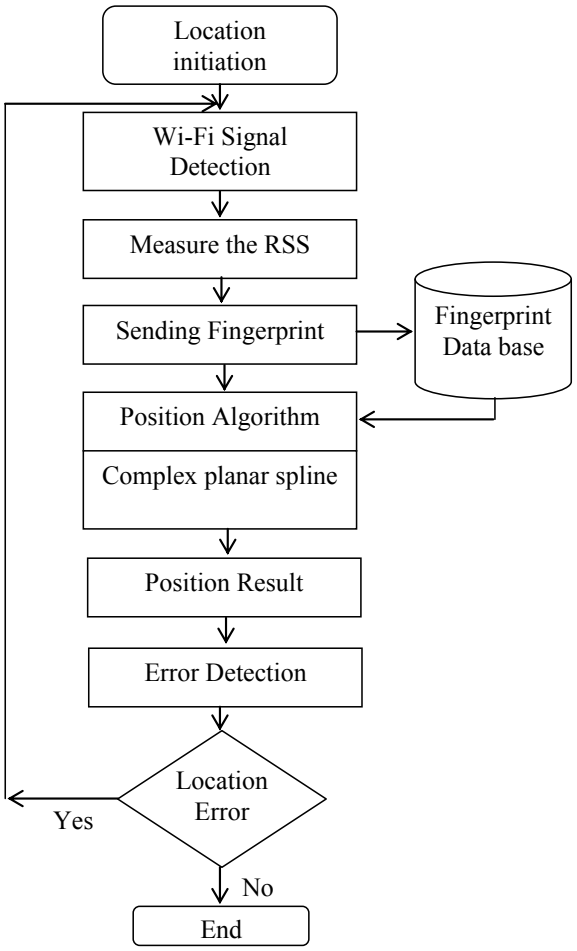



**Fig. 5** Algorithm for positioning the user in the Wi-Fi/Indoor network

## 3 Use of Linear Complex Planar Splines to Increase the Accuracy of User Locating in the Wi-Fi/indoor Network

Consider the process of constructing planar splines [19, 20]. Let $\overline{G} \subset Q$, where $\overline{G} = G \cup \partial G$, $\partial G$ is the boundary of the domain $G$, $Q = [a, a + H] \times [b, b + H]$ is a square with side $H > 0$ (Fig. 4), $N$ is a natural number, $h_N = \frac{H}{N}$, $x_k = a + kh_N$, $y_j = b + jh_N$, $k, j = \overline{0, N}$. Then $Q = \bigcup_{k,j=0}^{N-1} Q_{k,j}$, where $Q_{kj}$ is square cells with a pitch $h_N$, $Q_{k,j} = \{z = x + iy : x \in [x_k, x_{k+1}], y \in [y_j, x_{j+1}]\}$.

This partition is denoted by $\Delta_N$. We define $G_N$ as the union of all $Q_{k,j}$, for which $Q_{k,j} \cap G \neq \emptyset$. We divide each square $Q_{k,j} \subset G_N$ by the diagonal into two triangles $P^1_{k,j}$ and $P^2_{k,j}$. We denote such a partition by $\Delta'_N$. The index $N$ in $h_N$, $\Delta_N$, $\Delta_N{}'$ will be omitted if a fixed partition is considered. Let us denote by $G'_N$ the union of all $P^1_{k,j}$ and $P^2_{k,j}$, for which $P^1_{k,j} \cap G \neq \emptyset$, $P^2_{k,j} \cap G \neq \emptyset$ [14].

Consider one of the triangles $P^1_{k,j}$, $P^2_{k,j}$ included in $G'_N$ with such vertices $V_1$, $V_2$, $V_3$ that the conditions [14]:

$$\mathrm{Im}\, V_1 = \mathrm{Im} V_2, \mathrm{Re} V_2 = \mathrm{Re} V_3$$

From the triangulated domain $G'_N$, we construct a linear planar complex spline $S_\Delta(z)$, interpolating the function $f(z)$ at the vertices of the triangles, $P^1_{k,j}$, $P^2_{k,j}$, by setting [14]

$$S_\Delta(z) = a + bz + c\overline{z}, \tag{2}$$

where $z = x + iy$, $\overline{z} = x - iy$.

$$S_\Delta(z) = a + b(x + iy) + c(x - iy),$$

$$S_\Delta(z) = a + bx + cx + i(by - cy) = ReS_\Delta(z) + iImS_\Delta(z),$$

where $ReS_\Delta(z) = a + bx + cx$, $ImS_\Delta(z) = by - cy$.

Then, according to the interpolation condition at the points $z_{k,j} = x_k + iy_j$:

$$S_\Delta(z_{k,j}) = f(z_{k,j}),\ f(z) = Ref(z) + iImf(z),$$

$$ReS_\Delta(z) = Ref(z),\ ImS_\Delta(z) = Imf(z).$$

The values of the coefficients *a, b, c* of the linear planar complex spline $S_\Delta(z)$ are determined as follows:

$$a = \left\{f_1\left(V_2\overline{V_3} - \overline{V_2}V_3\right) + f_2\left(\overline{V_1}V_3 - V_1\overline{V_3}\right) + + f_3\left(V_1\overline{V_2} - \overline{V_1}V_2\right)\right\}/\delta, \tag{3}$$

$$b = \left\{f_1\left(\overline{V_2} - \overline{V_3}\right) + f_2\left(\overline{V_3} - \overline{V_1}\right) + f_3\left(\overline{V_1} - \overline{V_2}\right)\right\}/\delta, \tag{4}$$

$$c = \{f_1(V_3 - V_2) + f_2(V_1 - V_3) + f_3(V_2 - V_1)\}/\delta, \tag{5}$$

where $\delta = 2iIm\left(V_1\overline{V_2} + V_2\overline{V_3} + V_3\overline{V_1}\right)$.

For example, finding the coefficients of a linear planar complex spline $S_\Delta(z) = a + bz + c\overline{z}$ for a triangle with vertices $V_1 = 0$, $V_2 = 1$, $V_3 = i$ (Fig. 7), according to formulas (3–5), is defined as:

$$a = f_1,\ b = \tfrac{1}{2}\{f_1(i-1) + f_2 + if_3\},$$

$$c = \frac{1}{2}\{-f_1(i+1) + f_2 + if_3\},$$

where $\delta = -2i$, $f_1 = f(V_1) = S_\Delta(V_1)$, $f_2 = f(V_2) = S_\Delta(V_2)$, $f_3 = f(V_3) = S_\Delta(V_3)$.

Lemma 1. Let the function $f(z)$ is continuous in the cellular area $\overline{G_N}$, and $S_\Delta(z)$—be a complex planar spline of the form (1), interpolating the function $f(z)$ at the nodes $\{z_{k_{ij}}\}$. Then

$$|f(z) - S_\Delta(z)| \le 2\omega(f;h),$$

where $h = h_N$ and $\omega(f, h)$—modulus of continuity of function $f(z)$ in the area $\overline{G_N}$.

Proof. Similarly to [21], dividing $ReS_\Delta(z)$ and $ImS_\Delta(z)$, we obtain splines of the first degree of one variable:

$$ReS_\Delta(z) = a + bx + cx,\ ImS_\Delta(z) = by - cy,$$

interpolating functions $Ref(z)$ and $Imf(z)$ at the top of the squares $Q_{k_{ij}} \subset \overline{G_N}$, respectively.

Let be $z \in Q_{k_{ij}}$. As shown in [16], for the class of continuous functions, the following inequalities hold:

$$|Ref(z) - ReS_\Delta(z)| \le \omega(f;h),$$

$$|Imf(z) - ImS_\Delta(z)| \le \omega(f;h),$$

from which follows

$$|f(z) - S_\Delta(z)| \le 2\omega(f;h),$$

where $h = h_N$ and $\omega(f, h)$—modulus of continuity of function $f(z)$ in the area $\overline{G_N}$.

Consider using linear complex planar splines to position a user on a Wi-Fi/Indoor network using a modified Fingerprinted positioning method. Let us construct for the considered area shown in Figs. 4 and 6 linear complex planar spline (Fig. 7).

The input data for determining the user's coordinates using linear complex planar splines are shown in Table 1. For each user in the Wi-Fi/Indoor network the received power level values of the RSSI signal from the $AP_i$, $i = \overline{1,3}$ access points $RSSI_{1,1}$, $RSSI_{1,2}$, $RSSI_{1,3}$. Determining the user's location using the formula (1–5). The

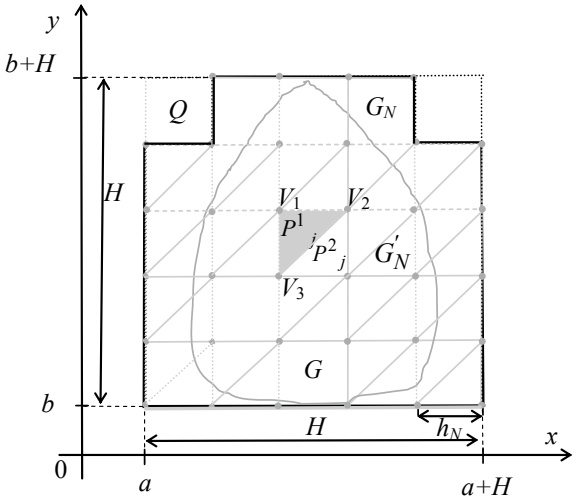



**Fig. 6** Construction of the mesh area $G$

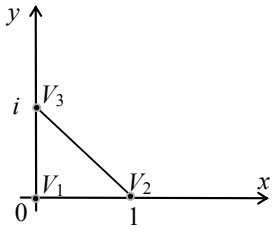



**Fig. 7** Element of triangulation with vertices $V_1 = 0$, $V_2 = 1$, $V_3 = i$

resulting RSS values made it possible to construct a linear complex planar spline for a given area G. Initial RSS, dBm values for determining user coordinates using linear complex plane splines are shown in Table 1 (Fig. 8).

Using a combination of Fingerprinted and finite element method based on a linear complex planar spline, user positioning accuracy can be improved by using quadratic or cubic complex planar splines.

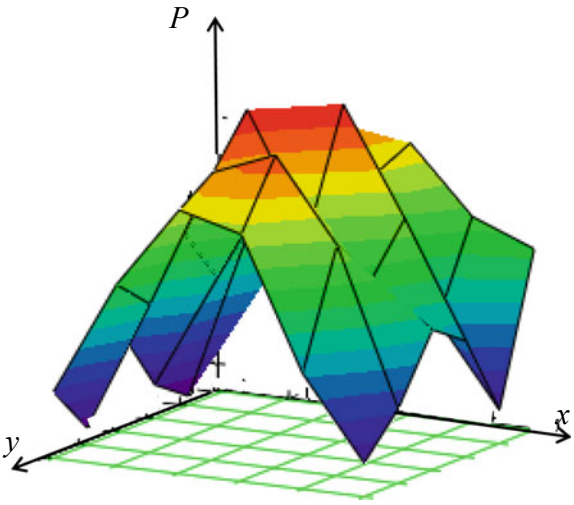



**Fig. 8** Construction of a linear spline function in the area G

**Table 1** Initial RSS, dBm values for determining user coordinates using linear complex plane splines

| *XY* | $x_0$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|---|
| $y_0$ | −38 | −35 | −44 | −50 | −42 | −39 |
| $y_1$ | −28 | −40 | −54 | −48 | −31 | −21 |
| $y_2$ | −31 | −35 | −49 | −58 | −45 | −35 |
| $y_3$ | −69 | −42 | −58 | −44 | −39 | −34 |
| $y_4$ | −39 | −35 | −52 | −55 | −45 | −46 |
| $y_5$ | −45 | −39 | −49 | −47 | −38 | −67 |

## 4 Conclusions

1. The analysis of existing methods for determining the user's location in the Wi-Fi/Indoor network is considered, the need for a new approach is shown, which would improve the accuracy of the user's location in the network.
2. To improve the accuracy of determining the user's location in the Wi-Fi/Indoor network, a modified method based on the Fingerprinting method and the finite element method using linear complex planar splines is proposed.
3. To solve the problem of determining the user's location in a Wi-Fi/Indoor network, it is proposed to use linear complex planar splines, and find their coefficients.
4. The direction of further research is the consideration of a modified Fingerprinting method based on complex planar splines (quadratic, cubic, B-splines,

etc.) and a comparative analysis of the results obtained to improve the accuracy of determining the location in the Wi-Fi/Indoor network.

## References


1. Liu KJR, Wang B (2019) Wireless sensing, positioning, IoT, and communications, 1st edn. Cambridge Univerity Press, Cambridge
2. Hameedah SH, Hussein M, Saad MS, Dzahir MA (2018) An overview of local positioning system: technologies, techniques and applications. Int J Eng Technol 7(3):1–5
3. Liao L, Lin K, Wang X, Lai X (2011) WiFi Positioning: a survey. Int J Commun Netw Distrib Syst 7(3–4):229–248
4. Gu F, Hu X, Ramezani M et al (2019) Indoor localization improved by spatial context - a survey. ACM Comput Surv 52(3):1–35. https://doi.org/10.1145/3322241
5. Sakpere W, Oshin MA, Mlitwa NBW (2017) A state-of-the-art survey of indoor positioning and navigation systems and technologies. South Afr Comput J 29(3):145–195. https://doi.org/10.18489/sacj.v29i3.452
6. Uryvsky L, Martynova K (2019) Complex analytical model of priority requires service on cloud server. In: 2019 international conference on information and telecommunication technologies and radio electronics (UkrMiCo-2019), Odessa, Ukraine,, 9–13 September 2019. https://doi.org/10.1109/UkrMiCo47782.2019.9165323
7. Yeremenko O, Tariki N, Hailan AM (2016) Fault-tolerant IP routing flow-based model. In: 2016 13th international conference on modern problems of radio engineering, telecommunications and computer science (TCSET), pp 655–657. https://doi.org/10.1109/TCSET.2016.7452143
8. Lemeshko O, Nevzorova O, Hailan A (2014) The increasing convergence of coordination procedure in the implementation of multipath hierarchical routing. In: 2014 first international scientific-practical conference problems of infocommunications science and technology, pp 45–48. https://doi.org/10.1109/INFOCOMMST.2014.6992294
9. Yeremenko O (2015) Enhanced flow-based model of multipath routing with overlapping by nodes paths. In: 2015 second international scientific-practical conference problems of infocommunications science and technology (PIC S&T), pp 42–45. https://doi.org/10.1109/INFOCOMMST.2015.7357264
10. Lemeshko O, Yeremenko O (2018) Linear optimization model of MPLS Traffic Engineering Fast ReRoute for link, node, and bandwidth protection. In: 2018 14th international conference on advanced trends in radioelecrtronics, telecommunications and computer engineering (TCSET), pp 1009–1013. https://doi.org/10.1109/TCSET.2018.8336365
11. Yang Q, Zheng S, Liu M (2019) Research on Wi-Fi indoor positioning in a smart exhibition hall based on received signal strength indication. Wirel Commun Netw. https://doi.org/10.1186/s13638-019-1601-3
12. Mehmood H, Tripathi NK, Tipdecho T (2010) Indoor positioning system using artificial neural network. J Comput Sci 6(10). https://doi.org/10.3844/jcssp.2010.1219.1225
13. Li G, Geng E, Ye Z, Zhu H (2014) An indoor positioning algorithm based on RSSI real-time correction. In: 14th IEEE international conference on signal processing (ICSP)
14. Strelkovskaya, I., Solovskaya, I., Strelkovska, J.: The use of linear complex planar splines to improve the accuracy of determining the location of the user in Wi-Fi/Indoor networks. In: 2021 international scientific-practical conference proceedings (PICS&T 2021), pp 1–4
15. Strelkovskaya I, Solovskaya I, Makoganiuk A (2019) Spline-extrapolation method in traffic forecasting in 5G networks. J Telecommun Inf Technol 3:8–16. https://doi.org/10.26636/jtit.2019.134719
16. Strelkovskaya I, Solovskaya I (2019) Using spline-extrapolation in the research of self-similar traffic characteristics. J Electr Eng 70(4):310–3016. https://doi.org/10.2478/jee-2019-0061

17. Strelkovskaya I, Solovskaya I, Makoganiuk A (2020) Forecasting 5G network multimedia traffic characteristics. In: 2020 IEEE 15th international conference on advanced trends in radioelectronics, telecommunications and computer engineering (TCSET), Slavske, Ukraine, 25–29 February 2020, pp 982–987. https://doi.org/10.1109/TCSET49122.2020.235585
18. Strelkovskaya I, Solovskaya I, Strelkovskaya J (2020) Spline-extrapolation of video traffic of IoT-devices based on various cubic splines. In: 2020 International scientific-practical conference proceedings (PICS&T 2020), Kharkiv, Ukraine, 6–9 October 2020, pp 243–248. https://doi.org/10.1109/PICST51311.2020.9467937
19. Belyi V, Strelkovskaya I (1988) Approximation of functions by analytic complex splines in domains quasiconformal boundary. Ukr Math J 40(5):563–568
20. Opfer G, Puri M (1981) Complex planar splines. J Approxim Theory 31(4):383–402
21. Zavyalov YS, Kvasov VI, Miroshnichenko VL (1981) Methods of spline-functions. Science, Moscow

# Principles of Building Modular Control Plane in Software-Defined Network




**Oleksander Romanov, Mykola Nesterenko, Volodymyr Mankivskyi, and Oleksandr Zhuk**

**Abstract** Existing networks are based on equipment that is a monolithic physical elements. Each element of the network has features that depend on the manufacturer. This does not allow telecom operators to make changes to equipment functions, to flexibly operate resources and quickly introduce new services. The use of SDN technology can significantly increase efficiency by centralizing management and virtualizing the functions of network elements. At the same time, operators have the opportunity to independently develop applications, which significantly speeds up the process of providing new services to users. The paper shows that when implementing SDN, it is necessary to ensure compatibility with existing networks using legacy management technologies based on operations support systems (OSS). It should be possible to include a person in the control loop - an operator who will take part in solving poorly structured network management tasks that are not amenable to automation. The tasks that the SDN network management plane should solve are described. The requirements for the architecture of ONOS are formulated, which is proposed to be built in the form of a modular, symmetrical, distributed system. A mathematical model is proposed for predicting the required SDN network resource using standard network indicators: incoming load, QoS, throughput. The model allows solving the problem of forming a load distribution plan with given QoS indicators.

O. Romanov (✉) · M. Nesterenko (✉) · V. Mankivskyi (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremoga Avenue 37, Kyiv 03056, Ukraine
e-mail: a_i_romanov@ukr.net

M. Nesterenko
e-mail: nikolaiy.nesterenko@gmail.com

V. Mankivskyi
e-mail: v.b.mankivskiy@gmail.com

O. Zhuk (✉)
Military Institute of Telecommunications and Information Technologies Named After Heroes of Kruty, Moskovska Street 45/1, Kyiv 01011, Ukraine
e-mail: radiosenter222@ukr.net

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_17

**Keywords** SDN architecture · Operations support systems · Coordinator · Network element · Open network operating system · Northbound APIs · Southbound APIs · NE resource

## 1 Introduction

Existing networks are a collection of equipment that was created on the basis of specific hardware and software platforms from various vendors. Therefore, the introduction of new modern services on existing networks requires replacement or modernization of the old equipment park [1–4]. This approach leads to the emergence of long cycles of design, purchase of the necessary equipment and commissioning. All this negatively affects the efficiency of providing new types of services to users.

However, telecommunications operators networks today are made up mostly of "monolithic physical" network elements, where the control, monitoring, administration and data transfer functions are performed by physical devices. Often, a telecom operator is forced to build its network using equipment from one manufacturer, since in this case it is easier to ensure the compatibility of network elements and carry out upgrades.

The deployment of new services and services requires modification (replacement) of equipment. In this case, it is necessary to make the corresponding changes in turn on each network element. This approach makes the operator's network inflexible, complicates the introduction of new services and functions, and also increases the operator's dependence on specific vendor solutions. Therefore, at present, the issues of building networks based on the concept of Software Defined Networks (SDN) are in the center of attention of representatives of research organizations, universities and mobile operators. The largest contribution to the development of this area is made by representatives of the Open Networking Foundation (ONF) consortium, whose main activity is to accelerate the process of practical implementation of SDN and NVF [5].

There are several critical areas where SDN technology can achieve a high network effect:

1. Centralize control over the SDN network. A feature of the control system is that it is a logically centralized system that provides full control and intelligent management of network resources. The traditional network management methods that are common today work using a limited amount of information about the state of the network elements. Thanks to the functionally centralized management of resource utilization, restoration of operability, security level and service policies, it is possible to achieve a high level of optimality of the resulting solutions. This is due to the fact that the control system will solve problems based on complete information about the network resources and use more optimal methods for solving.

2. Virtualization of network elements, applications and services. A distinctive feature of SDN technology is that it allows you to abstract from basic technologies, equipment and network management. Algorithms for serving "Physical Network Elements" and "Virtual Network Ele Cons" will be the same. This will allow the operator to gain independence from the equipment supplier. He will be able to solve issues of comprehensive use of different type equipment at the level of interaction of physical and virtual elements.
3. Separation of functions and tasks of hardware and software. At the same time, as hardware tasks, it is possible to recommend the execution of functions that do not depend on the type of technologies used. This can be, for example, the power system, cooling system, processor performance and the amount of allocated memory. As the tasks of the software, it is possible to recommend the implementation of protocol functions that ensure the implementation of any new technology, the implementation of interfaces for the interaction of system elements, the creation of models of the elements of a controlled network. Then the programmability of the SDN network elements will allow you to control the behavior of the network using software outside the network devices that provide the physical connection. As a result, network operators will be able to adapt the behavior of their networks for the introduction of new services and customer support. By decoupling hardware from software, operators can quickly implement innovative differentiated services without being constrained by proprietary and proprietary platforms.
4. Openness of the SDN architecture. It should become a completely open system that provides the ability to interact with different types of equipment and different technologies. At the same time, network elements and applications must interact with each other through open APIs, and not through control interfaces that are closely related to equipment. Open APIs must support a wide variety of applications, including cloud orchestration, OSS/BSS, SaaS, and business-critical networking applications. In addition, intelligent software will be able to control equipment from multiple vendors with open programming interfaces. Finally, within SDN, intelligent network services and applications can run in a common software environment.
5. Providing opportunities for network operators to write their programs as an addition to the underlying software. This will allow applications to control the behavior of the network. Hence, SDN will allow users to develop their network applications, intelligently monitor network health, and automatically adapt network configuration as needed.

In order to ensure the effective use of the advantages of SDN, it is necessary to solve a number of problems, the main of which are:

- Analysis of the SDN architecture and definition of the functions of its elements in the process of solving control problems.
- Determination of the order of interaction of system elements using standardized API-interfaces. These APIs should provide access to attributes and mechanisms that describe the subject area and the specificity of equipment from various

vendors. At the same time, the API itself should be abstracted from the specifics of the supplier's equipment or technology.
- Determination of methods for constructing network models, allowing to reflect the main processes and performance indicators.
- Presentation of the main list of management tasks and their description in a formalized form.

## 2 SDN Architecture and Functions of Elements in Process of Solving Control Problems

Let's consider the SDN architecture and analyze the functions of its elements in the process of servicing requests from incoming requests from users. At the same time, we will focus on the elements that are directly involved in the process of managing the load maintenance. It should be noted that in the near future SDN networks will operate in the environment of traditional networks. This means that these networks will use outdated protocols and technologies. Therefore, the system architecture should include elements that will ensure the interaction of SDN networks with networks using outdated technologies in automatic and semi-automatic modes of operation. In addition, it should be possible to include a human operator in the control loop for manual input of commands related to the establishment of service policies, the assignment of priorities in the allocation of network resources and other tasks that do not have a formalized description. The architecture of such a system and its constituent components are shown in Fig. 1. This architecture is proposed for use in recommendations ONF TR-502 [6].

An SDN architecture may include the following main components:

- Data plane.
- Controller plane.
- Application plane.
- Management plane.

Consider the purpose and composition of the elements of the SDN architecture levels:

1. Data plane. It can contain one or more NEs. Each element is a set of resources for processing and forwarding traffic. Both physical network elements and virtual ones can act as resources. By resources we mean an abstract representation of the capabilities (performance) of basic network elements.
2. Plane of applications (services). It can contain one or some applications. Applications can be of various types. First, these applications are designed to provide SDN network architecture clinics. Such applications do not participate in the management process. They are a source of increasing the number of users on the network. As a result, increasing the operator's income. Secondly, these applications are designed to solve individual specific management tasks. These applications that have a modular horizontal structure and their number increases with new

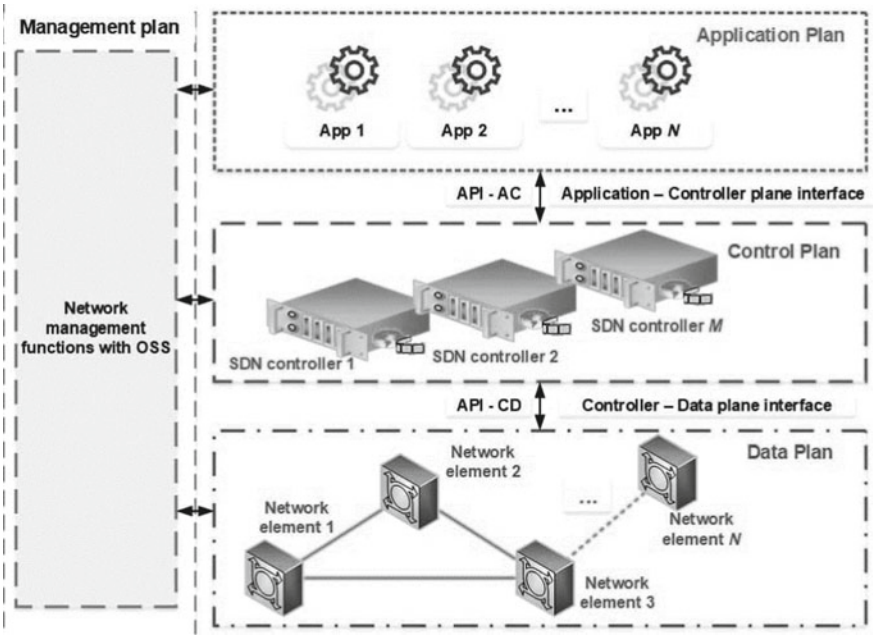



**Fig. 1** SDN with function management OSS (SDN с функцией management (OSS)

methods for solving management tasks. These applications are directly involved in the management process and are launched by the SDN controller as needed. Thirdly, applications that can act as an SDN controller. Such applications can take on some of the control functions. Can be used as a backup for the main SDN controller. But they should always work under the control of the SDN controller, which is assigned the main (leader).

Each application has its own set of resources over which it has exclusive control. However, this resource set can only be used by an application under the control of one or more SDN controllers. Applications can:

- Form requests for the use of network resources to the controller plane.
- Interact with each other using the SDN controller as a governing body.
- Directly interact with each other directly.
- Submit network requirements to the controller plane in the process of solving control problems.
- Act as an SDN controller.

3. Management plane. It may contain one or more SDN controllers. Each SDN controller is the authority for a set of network resources of one or more elements in the data plane or the application plane. The functionality of the SDN controller should provide the following network tasks:

- Collection of initial data for solving management problems.

- Control of the state of the main network elements.
- Processing requests for applications that are in the area of responsibility of its domain and are assigned to it.
- Isolation of applications from each other.
- Ensuring the interaction of applications.
- Control of network operation parameters.
- Access of a person - an operator with the required priority level for manual input of control commands.
- Timely response of the control plane to changes in the structure, load and quality of service in the network.
- Providing a solution to control problems to restore normal operation after a failure.
- Ensuring the service of incoming applications, taking into account the specified requirements.

To ensure the interaction of the control plane with the data and application planes, open interfaces are used. So for interaction with the data plane, the Application controller plan interface (A-CPI) is used. And to interact with the application plane, use

Data controller plan interface (D-CPI). The principle of construction and structure of these interfaces are in constant development. The interfaces currently most used will be discussed below, when considering

Open Network Operating System (ONOS), the main functional block of the SDN controller.

An important element of the SDN architecture is the management plan functional block. This block ensures the interaction of SDN networks with networks that incorporate outdated control systems built on the basis of Operations support systems (OSS). In addition, this functional block provides for the possibility of human intervention in the control process - the operator. Typically, it can be used for:

- Ensuring the interaction of the management systems to support the operations of the OSS networks of the old fleet with the SDN controller.
- Connection of a human operator to the SDN controller.
- Initial configuration of network elements.
- Assignment of network segments to SDN controllers, which will be under their control.

It should be noted that the initial setup of the network management system without the participation of a human operator is practically impossible. At this stage, there is a fairly large amount of poorly structured tasks that are not amenable to automation. In addition, there are management tasks that can only be solved by organizational measures. Therefore, during the initial launch of the SDN controller, the following tasks can be assigned to the human operator:

- Setting up user service policies.
- Defining the control and management areas that are provided for SDN apps.
- Introduction of restrictions on the use of the available resource.

- Determination of the list of system parameters to be controlled.
- At the application level, the human operator typically configures:
  - Contracts and service level agreements (SLAs).
  - Algorithms for solving control problems.
  - Device priority level in joint problem solving.

At all levels of the SDN architecture, the human operator sets up security policies that allow distributed functions to communicate securely with each other.

## 3 Functional Structure of the Modular Control Plane

Management of a system of any level of complexity involves collecting information about the state of network elements, identifying deviations of system parameters from normalized values, making decisions on methods to bring network characteristics to the required standards and forming control commands for executive bodies. To solve these problems, the control loop must include elements that are assigned these functions.

In accordance with the ONF TR-502 recommendations, the following functional blocks can be distinguished in the SDN architecture:

- Agents.
- Coordinators.
- Virtualizers.
- Resource databases (RDB).
- Managers.

The work [6] considers the generalized structure of such a system, which is shown in Fig. 2. Analysis shows that this structure requires clarification of the locations of functional elements, their purpose and functions, as well as the order of interaction in the process of solving control problems.

Let's start by looking at the functions of the "Agent" block. Usually this is the control object. The standard purpose of this function block is as follows:

- Monitoring the occurrence of failures in network elements located in the agent's area of responsibility.
- Collection of data on the performance of controlled network elements.
- Collection of statistical data on the network operation (port load, queue length, service delays).
- Collection of initial data for solving problems of optimizing the parameters of the network operation.
- Determination of deviations of controlled parameters from normalized values.
- Timely transfer of information to the control body about changes in the parameters of the controlled object.

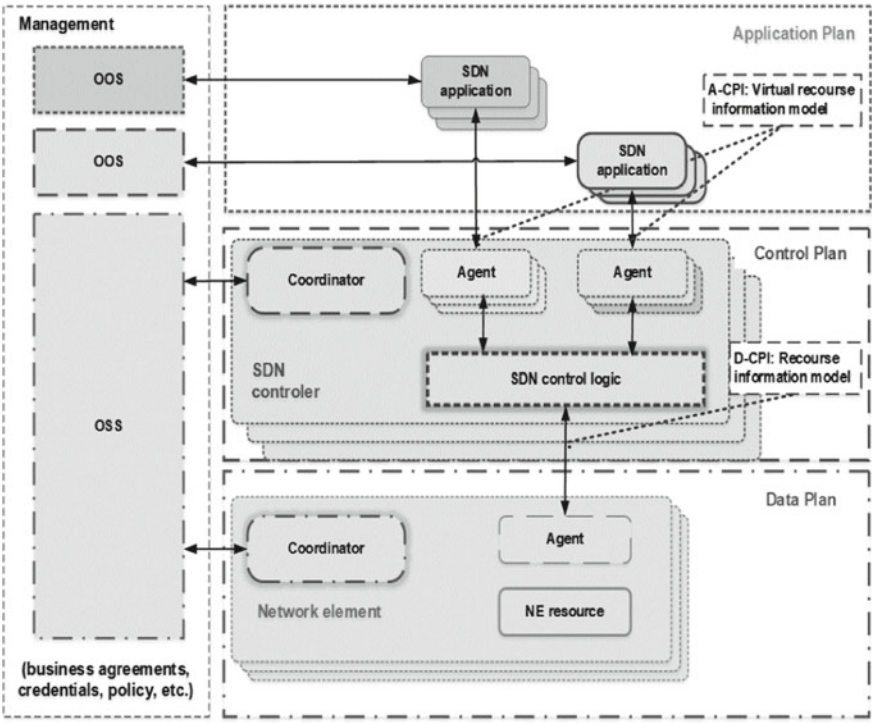



**Fig. 2** ONF TR-502 guidelines for placement of agents and coordinators in an SDN architecture

It should be noted that depending on the location of the agent, the list of tasks assigned to it may vary significantly. Therefore, you should immediately detail the location of the agent by assigning the appropriate index. Accordingly, to clarify the list of tasks to be solved by it. Based on the SDN architecture for the placement of agents and coordinators in Fig. 2, at least two groups can be distinguished:

- Agents that are located at the application layer and designate them as Agen

$$A_{AP\,i}, i \in (\overline{1, N}). \tag{1}$$

- Agents that are located on the Data plane level and designate them as an Agent

$$A_{DP\,l}, l \in (\overline{1, N}). \tag{2}$$

An important feature of SDN networks is that agents must control the network at different levels of abstraction. At the same time, the quality and level of network control should not depend on what type of elements the agent serves. It can be both physical and virtual elements. Therefore, it is advisable that the agent includes a model of the controlled object in its composition. In this case, the process of servicing

both virtual and physical network resources will be the same and will not depend on the equipment manufacturer.

In Fig. 3. The proposed functional structure of agents, recommended places for their placement and the order of interaction in the process of solving control problems are presented. In addition, here is a detailed diagram of the ONOS functional block, the location of coordinators, virtualizes and other elements.

The coordinator is a functional component, with the help of which a person—an operator can influence the order of the network functioning. It acts on behalf of a human operator and has the ability to provide and modify network service policy requirements. The coordinator can transmit control information to network elements in the form of data models from the control plane and the application plane. Therefore, the functional blocks of the coordinator are located everywhere.

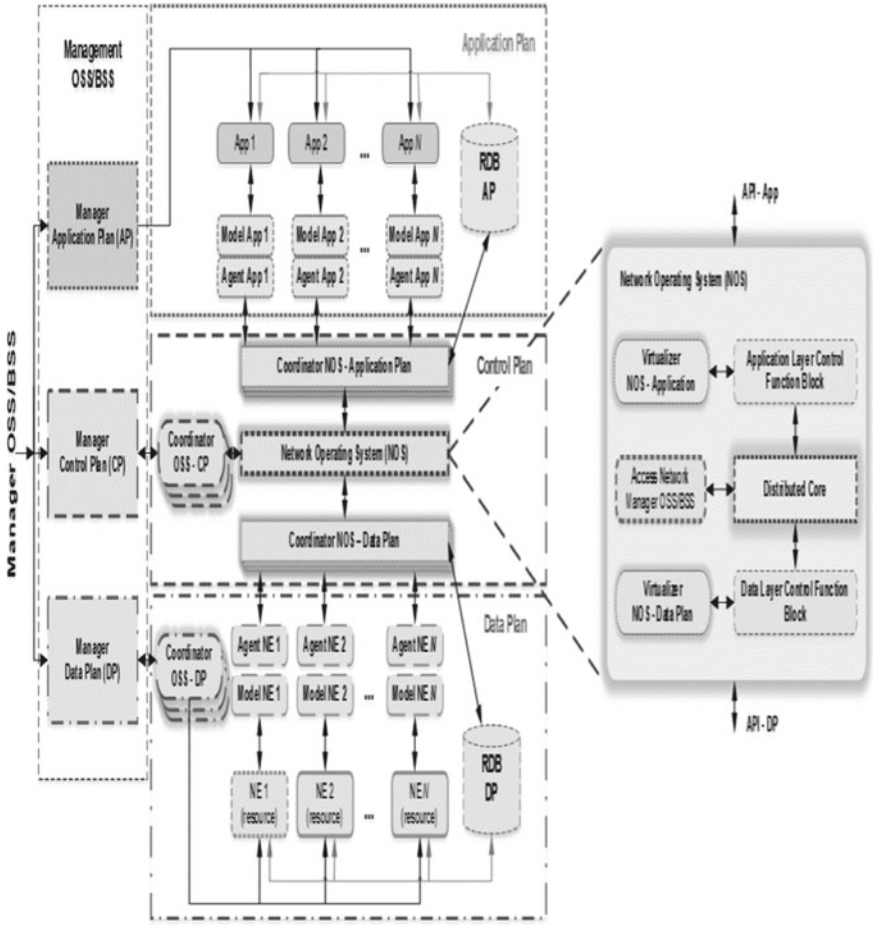



**Fig. 3** Functional structure of the modular control plane

An important function of the coordinators is to ensure the interaction of the SDN control layer with networks of legacy technologies that use OSS/BSS control systems. All coordinators are directly connected to the OSS/BSS Management level, that is, the level that can transmit control commands from a human operator with a high priority level.

There are four types of coordinators:

- The coordinators that are located at the Control plan level are responsible for the interaction of NOS and Application plan, and designate them as the NOS coordinator and Application plan:

$$C_{NOS_AP\ b}, b \in (\overline{1,\ N}). \tag{3}$$

- Coordinators that are located at the Control plane level are responsible for the interaction of the OSS/BSS and the Control plane, and designate them as a coordinator Control plane OSS/BSS:

$$C_{OSS/BSS_CP\ d}, d \in (\overline{1,\ N}). \tag{4}$$

- The coordinators that are located at the Control plan level are responsible for the interaction of NOS and Data plan, and designate them as the NOS Ap-plication plan coordinator:

$$C_{NOS_DP\ w}, w \in (\overline{1,\ N}). \tag{5}$$

- The coordinators that are located at the Data plane level are responsible for the interaction of the OSS/BSS and the Data plan, and designate them as the Data plane OSS/BSS Coordinator:

$$C_{OSS/BSS_DP\ y}, y \in (\overline{1,\ N}). \tag{6}$$

The main contribution to the solution of control problems is made by the coordinators of the Control plane level.

An important element of the control plane is the virtualizer. In the SDN architecture, virtualization solves the problem of forming and allocating virtual abstract resources for specific applications. The SDN controller offers application services as an instance of an information model that abstractly describes the available resources and the policy for their use.

A virtualizer is a functional object that stores an instance of the resource information model and policy for their use.

The virtualizer is created by the OSS/coordinator for each client application. The OSS/Coordinator allocates resources and defines the policy that the virtualizer should use when providing services to applications through the API. Next, the virtualizer creates an agent, into which its own copy of the information model of available

resources and the policy of its use is written to provide a service for a specific program.

Management in such a complex system as SDN is impossible without monitoring the parameters of the system functioning and the state of its elements. These data are used as initial data in solving control problems. To do this, the state of the system elements is monitored both at the Data plane and Application plan levels. The monitoring data is stored in a Distributed Resource Database (RDB). Based on these data, models of virtual network resources are created and the tasks of re-forming the load distribution plan are solved in cases of deviation of the network parameters from the required norms.

## 4 Open Network Operating System Architecture

### *4.1 Principles of Construction of ONOS*

Consider the principles of building an architecture, an open network operating system (ONOS) and its features. It should be noted that ONOS is the main functional block of the SDN controller. Therefore, in the scientific and technical literature, ONOS and the SDN network controller are often identified. It should be borne in mind that this is not entirely true. The controller consists of a number of functional blocks, one of which is ONOS. However, due to the fact that ONOS is entrusted with the solution of all intelligent network management tasks, its architecture uses approaches, descriptions and levels that were used both in describing the SDN architecture and in describing the controller of this network.

It should be noted that ONOS development is a separate open source ONF project distributed under the Apache 2.0 license [7]. In addition, the system description of ONOS is published in the public domain [8].

According to the ONF, the ONOS platform is being developed to meet the needs of operators to create carrier-level solutions that provide the flexibility to create and deploy new dynamic network services with simplified programming interfaces. ONOS should provide support for both configuration and real-time network management. This approach will eliminate the need to run routing and switching control protocols within the operator's network infrastructure.

Currently, a search is being made for possible ways to optimize and improve the efficiency of ONOS functioning. One of these ways can be the transfer of the ONOS functional block to the cloud. This will allow end users to create new network applications without the need to change the data plane.

A possible version of the ONOS architecture is shown in Fig. 4. As can be seen from Fig. 4. ONOS architecture design is modular. This approach allows you to ensure the optimal consumption of resources, since for each specific deployment, the optimal number of the required subset of modules will be determined.

ONOS architecture includes:

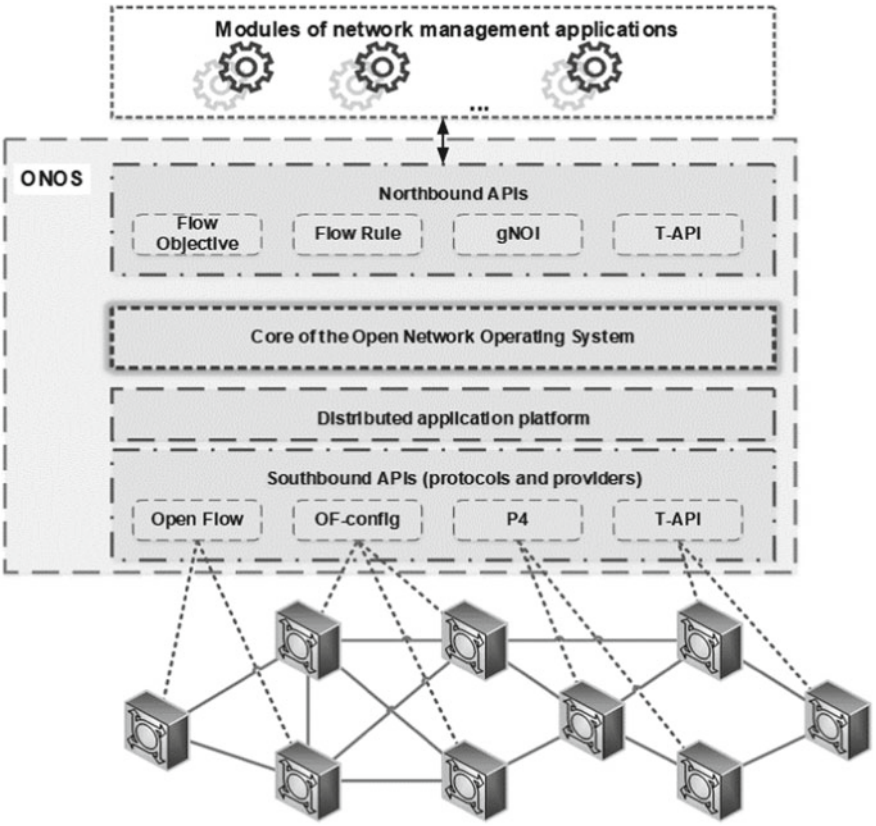



**Fig. 4** ONOS architecture

- The core of an open operating system, which is responsible for solving all the intellectual tasks of network management.
- Platform of distributed applications, which is a set of applications for solving network management problems. This platform is built as an extensible modular system. In addition, each modular functional block solves a specific control task. As the list of management tasks to be solved increases, the number of modules in the distributed application platform will increase.
- A set of open network interfaces. At the same time, just like in the SDN architecture, it is proposed to use two types of interfaces:
  - Open southern interface (SBI), which is a set of modules for interacting with the data layer.
  - Open northern interface (NBI), which is a set of modules for interacting with functional blocks of the application layer.

It should be noted that the south interface includes a set of protocols and interfaces, such as OpenFlow, OF-CONFIG, P4, T-API and others. The tasks that these elements of the system solve and their importance are determined by the conditions of the network functioning. For example, consider and compare the operation of the OF-CONFIG and OpenFlow protocols.

The OF-CONFIG protocol (OpenFlow Management and Configuration Protocol) is a protocol for configuring and managing the operational context of OpenFlow switches.

And the OpenFlow protocol is a protocol for managing the data processing process in routers and switches that implement software-defined network technology.

The OpenFlow protocol assumes that an OpenFlow switch (for example, an Ethernet switch that supports the OpenFlow protocol) has been configured with various parameters such as the IP address of the OpenFlow controllers and other network elements. That is, the switch is preconfigured.

The OF-CONFIG protocol is intended for remote configuration of OpenFlow switches that have not yet been put into operation. It works in a slower mode compared to the OpenFlow protocol.

For example, the OF-CONFIG protocol solves the problem of creating routing matrices, which will later be used by the OpenFlow protocol in real time when processing incoming packets. Another example of how the OF-CONFIG protocol works is to enable or disable a port. This feature is also not related to real-time packet processing.

OF-CONFIG represents an OpenFlow switch as an abstraction called a logical or virtual OpenFlow switch. The OF-CONFIG protocol allows configuration of the OpenFlow logical switch parameters so that the SDN controller can communicate with and control the OpenFlow logical switch via the OpenFlow protocol.

The OF-CONFIG protocol allows you to dynamically associate resources with specific OpenFlow logical switches. This protocol can make several virtual switches from one physical switch and assign a certain resource to each of them. ONOS uses these protocols to interact with the data plane in order to implement network management tasks.

Recently, P4, T-API are starting to be used. They are increasingly used to build network models in solving control problems. At the same time, P4 shows greater efficiency than OpenFlow.

It is advisable that the set of open network interfaces and ONOS protocols be universal for all network segments and elements. This would make it possible to build homogeneous, scalable, universal systems with a high level of interaction of all elements.

This ONOS architecture will allow:

- Simplify the process of reinstalling and upgrading software on network elements.
- Reduce the cost of introducing new technologies and services throughout the operator's network.
- Step-by-step increase in the number of control tasks to be solved by adding new functional blocks as part of the platform of distributed applications.

- Create a horizontally scalable system that provides a high level of fault tolerance, which is very important to ensure that the specified requirements for the reliability of the functioning of centralized control systems are met.

Based on the above, it is possible to formulate requirements for the architecture of ONOS. First, the horizontal plane of the SBI must be large enough. This is due to the fact that any access to the underlying equipment will be through ONOS. Therefore, the totality of all southern APIs should be sufficient to set up, maintain, and manage the network.

In addition, a multi-level system of applications and services running on ONOS should be provided. Depending on the priority of the task solved by the application and the frequency of its use, the level of efficiency of task processing should be determined. For example, applications that define the main control functional block in the event of failure of network elements should have a direct connection with ONOS. And applications that provide the installation of new software, certificates, configuration parameters and use the protocol.

Another requirement for ONOS is to provide simultaneous two-way exchange of information through NBI and SBI. As seen from Fig. 4, through NBI, applications use ONOS to manage network elements, and through SBI, southbound modules transmit information about the state of the core network to the ONOS core.

Interaction between the ONOS core and network devices is provided by a set of protocols and interfaces, such as OpenFlow, OF-CONFIG, P4, T-API, which take care of the details of interaction with devices, thereby isolating the ONOS core and applications running on top of it from the details of the variety of network devices.

The core of ONOS consists of a number of subsystems, each of which is responsible for a certain aspect of the functioning of the network. Each subsystem maintains its own service abstraction, which is responsible for propagating network state parameters across the cluster.

## *4.2 ONOS Subsystems and Services*

When building complex control systems, various methods are used. ONOS is built as a modular, symmetrical, distributed system. The principle of symmetry means that all blocks are functionally and programmatically of the same type. At the same time, one node, if necessary, can take over the performance of part of the functions of another node. The principle of modularity assumes that the operating system consists of separate subsystems of a modular type. To implement inter-domain interaction and scale fault-tolerant distributed systems, the Java reactive framework Atomix is used.

An important function of Atomix is the coordination of all ONOS instances. This accomplishes three tasks:

- Determination of the number of required instances of ONOS, which must function at any given time, depending on the workload and quality of service requirements. Solving this problem allows you to provide the required level of availability in

the event of failures. Atomix keeps track of healthy ONOS instances and failed ONOS instances.

- Performance of the main functions by each instance of ONOS to monitor network elements and maintain a subset of the physical switches of the cluster in which it is the leader.
- Timely determination of the ONOS leader in the group. The solution to this problem is carried out using the functional block (blocks) Atomix. It should be noted that all ONOS instances can monitor the status of switches and other network elements. But only the leader can manage network elements. If an ONOS instance fails, then Atomix ensures that a new leader is elected for switches and ONOS instances. The same approach is used if a new switch is connected to the network.

The ONOS application management subsystem takes responsibility for distributing application primitives throughout the cluster. It ensures that all nodes run the same software.

When solving management problems, ONOS uses standard protocols, interfaces and models, such as OpenFlow, NETCONF, OF-CONFIG and others. At the same time, an important property of ONOS is that its system architecture is not directly tied to them.

The base distribution of ONOS contains more than 175 modular applications (e.g., traffic management programs, device drivers, utilities, monitoring programs, ready-made YANG models). Consider the composition, purpose and principle of operation of the main ONOS modules (Fig. 5) [5].

In order to interact with the outside world, ONOS has a graphical interface and a number of external adapters such as a REST API, CLI, and an expanding dynamic web interface. It also uses an open source gRPC interface that uses HTTP/2 for transport. It provides features such as authentication, bi-directional streaming, flow control, cancellation, and timeouts.

The main intellectual component of ONOS, which is entrusted with solving most of the control tasks, is the ONOS core. It includes a set of applications that act as functional blocks of the kernel extension. These can be models of various network elements, sets of network protocols for implementing a particular technology, libraries of numerous drivers that solve the problems of interaction between functional blocks in the process of solving control problems.

Conventionally, the ONOS core can be divided into two parts: network and service (operational). The network part is responsible for solving intellectual problems related to ensuring the specified indicators of the network functioning. The network part supports the following functions: statistics collection, topology analysis, device configuration, creation of virtual networks, web groups, support for Internet services.

The service (operational) part of ONOS is assigned tasks related to servicing and maintaining the normal functioning of the hardware implementation of the SDN controller, as a separate physical element of the control plane. Examples of such tasks include device memory management, local configuration of the operating system, provision of messaging between controllers, support for a graphical interface for simplified system management.

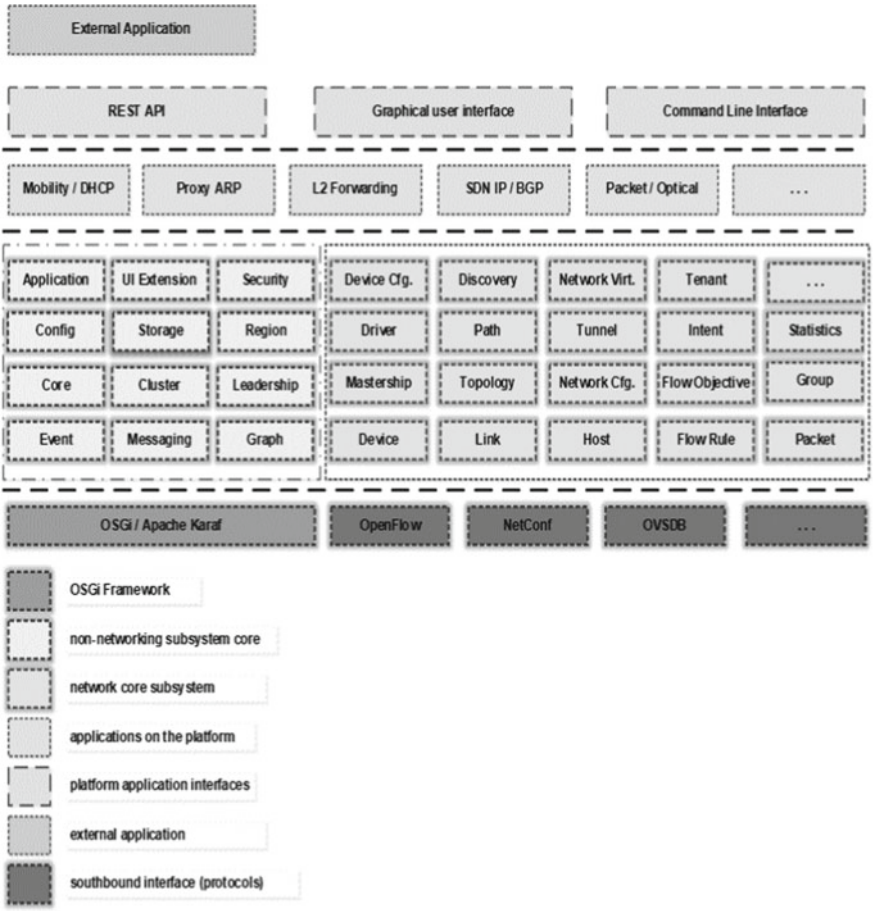



**Fig. 5** ONOS subsystems

As an example, we will describe the purpose of some ONOS network services, which are the most typical in solving network problems:

1. Mastership. Determines the leader using Atomix. Selects which ONOS instance in the cluster will act as a leader (master) for all infrastructure devices. In case of failure of any instance of ONOS, it ensures the timely selection of a new leader (master) for all remaining devices.
2. Cluster. Provides configuration management for an ONOS cluster. It provides information about all ONOS nodes in Atomix. The Atomix service performs clustering on ONOS nodes with the definition of a leader (master). At the same time, the ONOS nodes themselves are actually simple clients that are used to scale the control logic and enter information for network devices.

3. Network cfg. Provides a complete description of the network structure. Contains information about the list of network elements, their functionality, port load, host status, list of links. Provides information to external users about how the network is maintained by the ONOS core and its applications.
4. Config. Manages configuration settings for various software components in the ONOS core and applications. Operates with the following parameters:

   - Algorithm for processing the external stream.
   - Device address or DHCP server.
   - Polling frequency, etc.

This allows you to solve software configuration tasks. The parameters are set by the human operator in accordance with the accepted service policies.

5. Device. Stores information about infrastructure devices such as switches, routers, ROADMs. May include a description of the ports, their loading and delay time in the processing.
6. Host. Collects and stores information about the list and status of end systems connected to the network. This can be done using the ARP, NDP, or DHCP packet sniffing method.
7. Topology. Represents a description of a list of network elements. In its work, the service uses the Device and Link services. The network structure is described as a directed graph.
8. Packet. Provides interception of packets circulating in the network segment. It analyzes them, collects information and sends them back to the network. To solve these problems, widely known methods for discovering nodes and channels can be used, for example, ARP, DHCP, LLDP. This service is auxiliary to a number of basic services and applications. It is used by many applications because it provides information about network devices and their topology.

However, there are many more services. For example, some of them allow applications to program the behavior of the network using different levels of abstraction and different constructs. Such services can be:

- Path: defines the prefix or port to which the packet should be sent when performing the next step.
- Mcast: determines the IP address of the group, the location of the source and destination.
- Group: Groups ports or activities in a device, specifying a group of devices. This allows complex forwarding features such as cross-port load balancing, switching between ports in a group, and multicasting to all ports specified in the group. The group can also be used to aggregate the joint activities of different threads. At the same time, by changing one entry, we can change the order of service in a group of devices.
- Meter: can be used to limit the rate in order to provide a given quality of service for the selected network traffic.

- Flow Rule: used to program the packet forwarding algorithm in a pipeline-independent manner.
- Intent: Provides a topology- and network-structure-independent way of establishing flow rules. At the same time, the specifications may indicate various restrictions for using the end-to-end path, such as the type of traffic, the incoming and outgoing port, and the target host. The service maintains connectivity by continuously monitoring the network and changing routes to maintain performance in a changing network situation.

Each service has its own distributed database and notification capabilities. Some applications may extend this set with their own services.

Thus, ONOS is a convenient open source platform that allows different teams of developers to jointly participate in projects to modernize and improve the control system. At the same time, the use of ONOS allows you to build a logical centralized control plane in SDN networks. The existing set of functional modules, services and interfaces as part of ONOS allows you to perform network management tasks. For the further development of ONOS, it is necessary to develop mathematical models and methods for optimally solving control problems in different operating conditions, which will later become software modules at the application level.

## 5 Mathematical Model of ONOS Applications and Services

The considered principles of building the control plane of the SDN network have great potential in terms of the effectiveness of the decisions made. However, in order to realize these possibilities, it is necessary to develop mathematical models that adequately reflect the functioning processes and parameters of the controlled object. It is important that when building a model, standard parameters of the network functioning are used, which have a clear physical meaning for everyone.

These options can be:

1. Incoming load. In this case, it is necessary that the load in each information direction from host to host be taken into account.
2. Quality of service. Again, the quality of service needs to be evaluated in each information direction from host to host. In this case, an indicator of the quality of service, for example, may be the probability of refusal to service the application due to the lack of the required network resource.
3. Performance of branches (channels, virtual tunnels, transmission paths) between network elements. At the same time, it should be borne in mind that each branch is involved in servicing the load coming from several information directions. In networks with a high level of connectivity, each branch is involved in servicing the load that comes from 60% of information directions and more.

Consider an example of building a mathematical model for one of the control tasks. At the same time, I would like to note that the nature of the management tasks to be solved can be considered in various conditions:

1. Predicting the required performance resource of the branches at the stage of deploying the SDN network.
2. Management of the SDN network in the normal operation mode, when the load and quality of service in the network correspond to the normative values.
3. SDN network management under extreme conditions, when a number of network elements failed or there was a load surge, which led to a general network congestion.

For each of the above operating conditions, specific groups of control tasks can be distinguished. We will consider only the first group of management tasks, which is related to the prediction of the required performance resource of branches at the stage of deploying an SDN network. In this case, select the following tasks:

- Analysis of the topology (structure) of the network being deployed.
- Analysis of the existing (dedicated) network resource.
- Analysis of requirements for quality of service QoS and network capacity.
- Formation of a load distribution plan (PRN).
- Distribution of a network resource that ensures the implementation of the PRP with minimal time, material, technical and other types of costs.
- Formation of commands for network elements that implement the decisions made.

The tasks of this group can be assigned to ONOS services or assigned to applications. But given that this is a forecasting task, and it does not require a real-time solution, it is expedient to assign it to one of the applications.

The main task of forecasting is to determine the required performance of the branches, namely the necessary network resource that will provide service to the incoming load with a given quality.

The formalized statement of this problem can be written as follows:

1. Determine the required performance of each branch of the network, expressed by the number of virtual channels (tunnels):

$$V = \|V_m\|, \tag{7}$$

where $m = \overline{1, M}$, $M$—conditional number of the branch between switching nodes in the network), in the network given by the graph G (N, M).

2. Form a load distribution plan (PRN) set of paths (routes) $\mu_{ij}^{v}$, which is a matrix:

$$R_v = \left\|\mu_{ij}^{v}\right\|, \tag{8}$$

where $v = \overline{1, K}$, $v$—the order in which the route is used, and is the number of possible routes).

3. Providing service for the load entering the network; transfer of information between each pair of network elements (switching nodes) (outgoing node) and (incoming node)

$$Z = \left\| Z_{ij} \right\| \to ID_{ij}(i, j = \overline{1,\ N}). \tag{9}$$

where N—switching node number.

4. Providing information directions with a given quality of service with the minimum total network resource involved:

$$P = \left\| P_{ij} \right\| \ V_{\Sigma} \ (\| V_m \| \to \min), \tag{10}$$

branch Net $m$:

$$\| V_m \| = f\left(R_v = \left\| \mu_{ij}^{v} \right\|\right), \quad (m = \overline{1, M}; \qquad i, j = \overline{1, N};\ v = \overline{1, K}) \tag{11}$$

with the following restrictions:

$$\begin{cases} V_{\Sigma} \to \min \sum\limits_{m=1}^{M} V_m, \quad (m = \overline{1,\ M}) \\ P_{\text{norm}} \le \max \forall \left\| P_{ij} \right\|, \ (i, j = \overline{1,\ N}, \ \text{at}\, i \ne j) \\ P_{ij} = \prod\limits_{k=1}^{v} [1 - \prod\limits_{i=1}^{M} (1 - p_m)], \ (i, j = \overline{1,\ N}, \ \text{at}\, i \ne j; \ v = \overline{1,\ K}; \ m = \overline{1,\ M}) \\ Z_{ij} \le \sum\limits_{m=1}^{M} Z_m, \quad (i, j = \overline{1,\ N}, \ \text{at}\, i \ne j; \ m = \overline{1,\ M}) \\ U = \max \frac{1}{M} \sum\limits_{m=1}^{M} U_m, \ (m = \overline{1,\ M}) \\ N, M, V_m, U_m \in G(N, M) \\ ID = N(N-1) \ \ at \ \ ID_{ij} = ID_{ji} \end{cases} \tag{12}$$

where $P_{\text{norm}}$—tolerable rate of denial of service in the data direction, $P_{ij}$—real (calculated) probability of denial of service in information directions ($i, j = \overline{1, N}$, at $i \ne j$), $p_m$—probability of losses on the network branches, $Z_{ij}$—service load in information directions, $Z_m$—distributed operating load on the network branches, $U_m$—network resource usage factor in branches Net ($m = \overline{1, M}$), $M$—numbers of branches in the network.

Depending on the dimension of the network structure, the number of its elements and algorithms for distributing information, the complexity of solving the problem can vary significantly. For networks of large dimensions, it is difficult, and sometimes impossible, to obtain a solution to the problem in this form. Then, the problem statement is simplified and the problem of the following form is solved:

$$V = \| V_m \|, \quad (m = \overline{1, M}) \tag{13}$$

with the following restrictions:

$$\begin{cases} V_{\Sigma} \to \min \sum\limits_{m=1}^{M} V_m, \quad (m = \overline{1,\ M}) \\ R_v = \left\| \mu_{ij}^{v} \right\|, \quad (i, j = \overline{1,\ N}, \quad \text{at}\, i \neq j; \quad v = \overline{1,\ K};) \\ P_{\text{norm}} \leq \max \forall \left\| P_{ij} \right\|, \quad (i, j = \overline{1,\ N}, \quad \text{at}\, i \neq j) \\ P_{ij} = \prod\limits_{k=1}^{v} [1 - \prod\limits_{i=1}^{M} (1 - p_m)], \quad (i, j = \overline{1,\ N}, \quad \text{at}\, i \neq j; \quad v = \overline{1,\ K}; \quad m = \overline{1,\ M}) \\ Z_{ij} \leq \sum\limits_{m=1}^{M} Z_m, \quad (i, j = \overline{1,\ N}, \quad \text{at}\, i \neq j; \quad m = \overline{1,\ M}) \\ U = \max \frac{1}{M} \sum\limits_{m=1}^{M} U_m, \quad (m = \overline{1,\ M}) \\ N, M, V_m, U_m \in G(N, M) \\ ID = N(N-1) \ at \ ID_{ij} = ID_{ji} \end{cases} \tag{14}$$

The difference in the formulations of the problem lies in the fact that in (2) in the process of calculating the required performance of the network branches, the assumption is made that the load distribution plan is given. It should be borne in mind that such an assumption can have a significant impact on the accuracy of the resulting solution. To eliminate this shortcoming, it is possible to recommend an iterative procedure in order to consider various options for servicing the load in the conditions of using different options for the load distribution plan.

## 6 Conclusions

The paper analyzes the architecture and capabilities of the SDN network. The areas of use in which SDN technology allows to achieve a high degree of network flexibility for the rapid deployment of new services and the provision of modern services to users are noted. The main advantages of SDN networks are: centralization of management in the SDN network; virtualization of network elements, applications and services; separation of functions and tasks of hardware and software; the openness of the SDN architecture and the use of standardized interfaces, which allows network operators to develop their applications (modules) as an addition to the basic software.

The recommendations of the ONF on the construction of the control plane of the SDN network are considered. It is noted that in the near future SDN networks will operate in the environment of traditional networks. Therefore, the interaction of SDN networks with networks using legacy management technologies based on Operations Support Systems/Business Support Systems must be ensured. In addition, it should be possible to include a person in the control loop - an operator who will take part in solving poorly structured network management tasks that cannot be automated.

The tasks are detailed, the solution of which should be assigned to the control plane. The elements of the system are determined, which provide the collection of information about the state of the network and ensure the identification of deviations

of the network parameters from the normalized values. And also, the functions of the selected elements of the system, their location and the order of interaction in the process of solving control problems are detailed, which is reflected in the functional structure of the modular control plane.

The principles of building an open network operating system are considered. The requirements for the architecture of ONOS are formulated. It is proposed to build ONOS in the form of a modular, symmetrical, distributed system. It is shown that such an approach to building ONOS applications and services has great potential in terms of the effectiveness of decisions made. But to realize this potential, mathematical models are needed that adequately reflect the functioning processes and parameters of the controlled object.

A description of a mathematical model for solving one of the possible control problems is given. Its advantage is that it operates with standard parameters of the network functioning, which have a clear physical meaning to everyone. These parameters are: incoming load, quality of service, performance of network branches (required amount of network resource).

## References

1. ONF. Accelerating the Adoption of SDN & NFV. Issue 1, June 2016
2. ONF TR-502. SDN architecture, Issue 1, June 2014
3. ONF TR-508. Requirements Analysis for Transport OpenFlow/SDN. V1.0, 20 August 2014
4. ONF specification. OpenFlow Table Type Patterns, Version No. 1.0, 15 August 2014
5. IETF RFC 7426. Request for Comments: 7426, Category: Informational K. Pentikousis (eds.), ISSN: 2070-1721 EICT
6. ONF TS-022. Optical Transport Protocol Extensions, Version1.0, 15 March 2015
7. ONF TR-539. OpenFlow Controller Benchmarking Methodologies, Version 1.0, November 2016
8. ONF TS-029. MPLS-TP OpenFlow Protocol Extensions for SPTN, Version 1.0, 16 June 2017
9. ONOS. Security and Performance, Report No. 1, 19 September 2017
10. ONOS Security and Performance, Report No. 2, 2 November 2018
11. ONF TR-537. Negotiable Datapath Model and Table Type Pattern Signing, Version 1.0 (2016)
12. ONF TR-522. SDN Architecture for Transport Networks, March 2016
13. Cascone C, Vachuska TH, Davie B (2021) Software-defined networks: a systems approach, p 152
14. Phemius K, Bouet M, Leguay J (2013) DISCO: distributed Multi-domain SDN controllers, Thales Communications & Security, 29 August 2013
15. Lam J, Lee S-G, Lee H-J, Oktian YE (2016) Securing SDN southbound and data plane communication with IBC. Hindawi Publishing Corporation Mobile Information Systems. https://doi.org/10.1155/2016/1708970
16. Phemius K, Bouet M, Leguay J (2013) ONOS Intent Monitor and Reroute service: enabling plug & play routing logic, Thales Communications & Security, 29 August 2013
17. Comer D, Rastegarnia A (2019) Externalization of Packet Processing in Software Defined Networking, 11 January 2019
18. ONF TR-525. SDN Interoperability Event Technical Issues Report AppFest (2015)
19. Romanov O, Oryschuk M, Hordashnyk Y (2016) Computing of influence of stimulated Raman scattering in DWDM telecommunication systems. In: UkrMiCo, pp 199–209

20. Globa L, Skulysh M, Romanov O, Nesterenko M (2019) Quality control for mobile communication management services in hybrid environment. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technologies. UKRMICO 2018. LNEE, vol 560, pp 133–149. Springer, Cham. https://doi.org/10.1007/978-3-030-16770-7_4
21. Romanov O, Mankivskyi V (2019) Optimal traffic distribution based on the sectoral model of loading network elements. In: IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), pp 683–688, August 2019
22. Yeremenko O, Persikov A (2018) Secure routing in reliable net-works: proactive and reactive approach. In: Shakhovska N, Stepashko V (eds) Advances in Intelligent Systems and Computing II, CSIT 2017, AISC, vol 689, pp 631–655. Springer, Cham. https://doi.org/10.1007/978-3-319-70581-1_44
23. Romanov O, Nesterenko M, Veres L, Kamarali R, Saychenko I (2021) Methods for calculating the performance indicators of IP multimedia subsystem (IMS). In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technology and Systems. MCT 2019. LNNS, vol 152, pp 229–256. Springer, Cham. https://doi.org/10.1007/978-3-030-58359-0_13
24. Phemius P, Bouet M, Leguay J (2013) Distributed Multi-domain SDN Controllers. Thales Communications & Security, pp 198–209
25. Romanov O, Siemens E, Nesterenko M, Mankivskyi V (2021) Mathematical description of control problems in SDN networks. In: International Conference on Applied Innovations in IT (ICAIIT), pp 33–40. https://doi.org/10.25673/36582
26. Romanov O, Nesterenko M, Mankivskyi V (2021) The method of redistributing traffic in mobile network. In: Ageyev D, Radivilova T, Kryvinska N (eds) Data-Centric Business and Applications. Lecture Notes on Data Engineering and Communications Technologies, vol 69, pp 159–182. Springer, Cham. https://doi.org/10.1007/978-3-030-71892-3_7

# The Method of Using a Telecommunication Air Platform as a Flying Info-Communication Robots

**Oleksandr Lysenko, Valery Romaniuk, Anton Romaniuk, Valery Novikov, Valery Yavisya, and Ihor Sushyn**

**Abstract** The paper is devoted to the conducted research of effectiveness of method of improved monitoring data collection which are accumulated in the sensors of the wireless sensor network. Data collection is performed by the so-called info-communication robot under different initial conditions: network dimension, number of clusters, number of nodes in the cluster, options for constructing data collection methods, flight strategy over nodes in the cluster. The results of efficiency comparison of using the improved method with the existing method of direct data collection and the centroid method of data collection by time criteria are given: the duration of data collection and the duration of network operation. Carried out the analysis of four strategies of flight over cluster (only between collection points centers; flight over critical nodes; data transfer in points closer to the flight route; cooperative). The efficiency of the improved method of data collection from the main nodes of the clustered network was evaluated.

O. Lysenko (✉) · A. Romaniuk (✉) · V. Yavisya (✉) · I. Sushyn (✉)
Telecommunication Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Industrialnyi Lane 2, Kyiv 03056, Ukraine
e-mail: lysenko.a.i.1952@gmail.com

A. Romaniuk
e-mail: anton.romaniuk@gmail.com

V. Yavisya
e-mail: yavisya@bigmir.net

I. Sushyn
e-mail: rubin268@gmail.com

V. Romaniuk (✉)
Military Institute of Telecommunications and Informatization, Moscow Street 45/1, Kyiv 01011, Ukraine
e-mail: romval2016@gmail.com

V. Novikov (✉)
Telecommunication Systems Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Industrialnyi Lane 2, Kyiv 03056, Ukraine
e-mail: novikov1967@ukr.net

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_18

**Keywords** Wireless sensor network · Telecommunication aero platform · Info-communication robot · Monitoring · Data collection · Clustering

## 1 Introduction

The classic telecommunication aero platform (TAP) is a base station, which is in three-dimensional space over the underlying surface due to the use of aircraft ("air cell") or aerostatic, or rotorcraft, or aircraft type [1, 2]. If real-time information processing is performed on the TAP board and decision is made basis on this processing to perform complex spatial movements ("intelligence of action"), which are aimed at improving the functional tasks of the network served by the TAP, the telecommunications aero platform should be considered flying info-communication robot (FICR).

The wireless sensor network (WSN) is considered as a network served by FICR. The task of FICR is collecting information (data) from the WSN. When using FICR the existing trivial method of data collection (the method of direct data collection from each sensor node WSN separately [2, 7]) does not allow to use all technological capabilities of FICR to increase the efficiency of WSN (increasing "lifetime" and/or reducing time of information collection—increasing the reliability, functional stability and survivability of WSN as the result). The main advantage of the trivial method is the simplified algorithms of functioning and interaction of WSN and FICR, which leads to reduce cost WSN hardware. But the importance of economic advantage concedes to reliability, functional stability and survivability of WSN in many critical situations [3–8].

The main idea of the development algorithms for the interaction of WSN and FICR is flexible (adaptive) clustering of WSN nodes, which provides an opportunity to implement a rational route of FICR movement. According to these algorithms, proposed to combine nodes into temporary clusters, where the role of the main node of the cluster is assigned to FICR. The algorithm for constructing the trajectory of FICR is based on known information about the coordinates of the position of the nodes WSN and calculates the coordinates of the data collection points, and then builds the trajectory of FICR ("intelligence of action"). In contrast to the existing "rigid" centroid clustering algorithms, in the article proposed to use "adaptive (flexible)" algorithms of cluster analysis, which are named respectively: k-average algorithm and formal element algorithm (for.el). These algorithms are characterized by less computational complexity and provide the ability to adaptively change the size of the cluster and, thus, control the quantity of clusters [4–7].

# 2 Main Part

## *2.1 Mathematical Problem Statement*

We consider parameters are invariant that characterizing:

1. technical and algorithmic properties of the network, nodes, FICR;
2. requirements for the quality of collecting information (data), requirements for the collecting monitoring information (data) system.

The control parameters are considered to be the quantity of points of the FICR traffic route (the quantity of points of information (data) collection that coincides with the quantity of clusters $n_k$) and the characteristics of the FICR traffic trajectory between these points and near these points; scheme of flight over route points of FICR.

The control parameters affect the criteria selected as: time of information collection (data) $T_{\text{coll}}$; time of functioning ("life") WSN $T_f$; quantity of FICR which selected for servicing WSN $N_{IAFICR}$.

Mathematical problem statement takes the form:

$$T_{\text{coll}} \rightarrow \min \text{ or } T_{\text{coll}} \leq T_{\text{collset}}, \tag{1}$$

$$T_f \rightarrow \min \text{ або } T_f \geq T_{\text{fset}} \tag{2}$$

$$N_{IAFICR} \rightarrow \min \text{ або } N_{IAFICR} \leq N_{IAFICRset} \tag{3}$$

where $T_{\text{coll}} = L_{\text{rb}}/v$;

$$L_{\text{rr}} = f((x, y)_k, h, t_{\text{flyk}}, St,\ k = 1 \ldots n_k), \tag{4}$$

when the restrictions $\Omega$ are the next:

1. he limit time (distance) in the round of flight IAFICR

$$T_{\text{coll}} \leq T_{\text{flymax}} (0 < L_{\text{rb}} \leq L_{\text{rr}} \leq L_{\text{rmax}}); \tag{5}$$

2. flight speed

$$\text{TAP} - v = [v_{\min}, v_{\max}]; \tag{6}$$

3. the quantity of clusters—$1 \leq k \leq n_k$;
4. battery energy of nodes and FICR—$e_i \leq e_{\max}$, $e_{\text{IAFICR}} \leq e_{\text{IAFICRmax}}$;
5. the coordinates of the location $(x, y)$ of the nodes are in the coverage area of the monitoring zone A;

6. flight altitude—$h = [h_{\min}, h_{\max}]$;
7. volumes of buffers $V_{\text{bufnode}i} \leq V_{\text{bufnodemax}}$; $V_{buf\text{IAFICR}} \leq V_{buf\text{IAFICRmax}}$;
8. requirements of service models, for example, with guaranteed quality of service, the flight time of each $k$-th cluster $t_{\text{flyclust}k}$ must be greater than the total transmission time between all nodes and TAP—$t_{\text{flyclust}k} \geq t_{\text{tr}k}$;
9. $St$—a set of strategies (schemes) of flight.

## 2.2 Solution Synthesis by Intellectual Adaptive Flying Info-Communication Robot

In the synthesis of the solution on board FICR regarding the choice of a rational method of collecting information from the nodes of the WSN uses a systematic approach (the components of which are presented in Fig. 1).

The system approach allows:

1. Provide intellectual and adaptability for algorithms of functioning of FICR.
2. Formulate the stages (components) of the method of collecting information from the nodes of WSN.

Consider the components of the method of collecting information from the nodes of the WSN in detail:

1. The method of direct data collection from each node of the network - with the flight over each node, with the flight over territory (where the WSN is located), with the flight over clusters of the WSN.
2. The method of optimizing operations regarding to management - isolated (control center of network, control system of FICR, control system of node WSN) or cooperative - in interaction between them; in the presence or absence of information about the state of nodes WSN in FICR (CC of network).

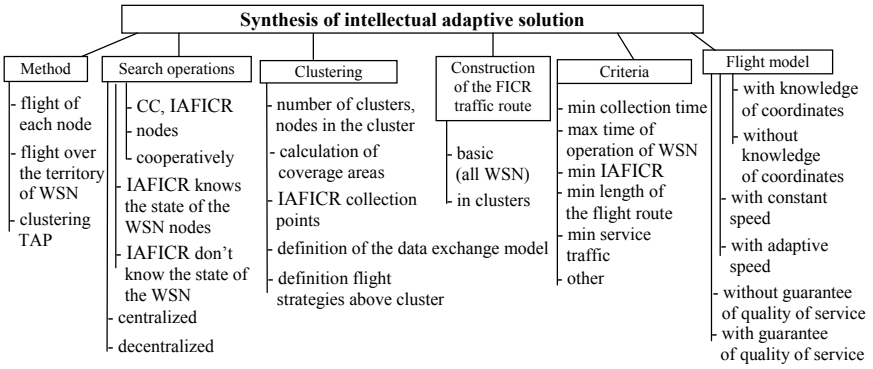



**Fig. 1** The structure of the system approach regarding to the choice of a rational way of collecting information from the nodes of the WSN

3. Procedure and rules of clustering WSN:
   - to calculate quantity and size of clusters (optimization of the quantity and size of clusters should be carried out depending on the various target functions of network control at the stage of data collection);
   - to cluster network and determine the points of collection monitoring data for certain parameters depending on the target functions of network control and the situation in the network;
   - to determine models (algorithms) of data exchange between FICR and cluster nodes;
   - to calculate strategies and flight parameters of FICR over nodes in clusters for given target functions (construction, optimization and adjustment the path of FICR flight over nodes depending on the target control functions and available resources of FICR and WSN nodes). If there is information about the status of the WSN, these tasks will solve centralized by control center of network (CCN). If it absent, these tasks will solve decentralized by interaction IAFICR and network nodes.
4. To construct FICR flight routes over data collection points: basic (flight over the whole WSN network) and flight routes over each cluster. For example, the basic flight parameters are calculated by CCN, and FICR adjusts the basic route in clusters after receiving information about the state of the cluster nodes when approaching it and the availability of appropriate resources.
5. Determine the criteria and their priority in managing the data collection process, namely: minimization of data collection time, maximum operating time of WSN, minimum FICRs quantity, minimization of energy consumption of nodes in the process of transmission from nodes to FICR, etc.
6. Determine the flight model based on the presence or absence of information about the coordinates of the WSN nodes, with a constant or adaptive speed, with a guarantee of service or its absence.
7. FICR flight models when collecting monitoring data from WSN. The formation and implementation FICR flight parameters take place in two stages.

In the first stage the control center of network builds the basic shortest route of the flight over points collection data in the WSN using one of the known algorithms solving the problem of the travelling salesman, (for example, the method of the nearest neighbor) determines the average speed and flight altitude.

In the second stage, when approaching the next cluster FICR adjusts the parameters of its flight (trajectory, speed, quality of service) depending on the quantity of nodes in the cluster, cluster node parameters (location, available node battery power and monitoring data volumes), available personal resources (energy and time remaining on the flight) and target network control functions. For example, if there

are "critical" nodes in the cluster ("exhausted", "overloaded", etc.), flight over them (servicing) is proposed to be carried out at a minimum distance to them with priority in servicing. To ensure the guarantee of data collection time, the required flight speed is calculated and implemented (FICR hangs at a certain point in space for the required time).

The average flight time (in general data collection) in the network $T_{\text{coll}}$ is determined by the ratio of the length of the flight route $L_{\text{r}}$ to the average flight speed $v$: $T_{\text{coll}} = L_{\text{r}}/v$. The maximum FICR flight time is determined by the simplest strategy data collection - directly from each node $T_{\text{collmax}} = L_{\text{rmax}}/v$, minimum time (with this clustering solution) is equal to $T_{\text{collmin}} = L_{\text{rmin}}/v$, where $L_{\text{rmin}}$—is the minimum route length covering all network collection points, provided that the time required to collect (transmit) data in each $k$-th cluster will be less than the time of its flight $t_{\text{tr}k} \leq t_{\text{fly}k}$ (with guaranteed servicing 'all monitoring data).

Reducing the length of the flight route reduces the time of information collection and reduces the energy consumption of the FICR, but increases the energy consumption of the sensor nodes due to the distance increase in the radio channels of the node-FICR.

Also, data collection time $T_{\text{coll}}$ depends on: application requirements, average flight speed $v$, quantity of formed network clusters $n_{\text{cl}}$, location $(x, y)_k$ data collection points, base route length $L_{\text{rb}}$ (real rout length $L_{\text{rr}}$ will be increases due to the application different flight strategies over clusters), flight altitude $h$, flight time of each $k$-th cluster of the network $t_{\text{flyclust}k}$, transmission speed in the radio channel node- FICR, flight strategy when servicing each k-cluster $St_k$, capabilities and resources of FICR.

It is advisable to consider different flight models:

1. Basic simplest model. Flight with the same constant speed in the cluster and between clusters simplifies the control of FICR movement, does not impose additional requirements on the UAV, can be used as a rotorcraft and aircraft type UAV. Data collection begins when radio communication is established with the first node of the cluster. The collection time is equal to the flight time through the cluster. If nodes did not have time to transmit data, it would store in the nodes until the next round of flight. Upon arrival in the radio communication zone with the base station, FICR transmits data to it and starts a new flight cycle. The model can be implemented as a rotorcraft and aircraft type UAV.
2. FICR flight with the same constant speed sufficient for cluster servicing and with increased speed (determined by the capabilities of the UAV) when moving between clusters. To implement the model requires UAVs with variable flight speed. This model can be implemented as a rotorcraft and aircraft type UAV.
3. Flight with adaptive speed in clusters. Certain types of applications may require different throughput, which are determined by external factors such as an emergency. For example, some nodes may have special capabilities for collecting and transmitting audio and video data, which may be required at certain times (the appearance of the attacker in the control area, high temperature, pressure and vibration in the pipeline, etc.). In addition, for guaranteed data collection, the throughput must be adapted to the required baud time. The decision to increase

or decrease the time of information collection is made by CCN or CS FICR together with the source node of this data. Rotorcraft and aircraft UAVs can be used.

4. Flight with guaranteed servicing nodes of cluster. FICR calculates the service time of all nodes at a constant throughput. Additional time is calculated in case of lack of time. It should be noted that this model can be used in emergency scenarios or special missions, where monitoring of a certain geographical area or "hot spot" is required for a certain period of time. The model can be used to transmit real-time traffic. Rotorcraft type UAVs are used.
5. Flight with a limitation of the maximum service time of the cluster (node). A node of a certain cluster is able to collect a fairly large amount of data, which requires a very significant amount of servicing time. At the same time other nodes then expect service in this flight cycle, the delay time in the servicing of certain applications may be exceeded, the buffers of other nodes of the cluster may be overflowed, and so on. Therefore, for fair service, a service time limit is set for each cluster. If the amount of data in the cluster cannot be serviced during this flight cycle, then the unserved portion of the data is carried over to the next round of flight. rotorcraft type UAVs are used. This model can also be used to ensure the security of the data collection system as a whole. It prevents FICR from capturing by enemy node that has requested an infinitely large buffer and organizes a "service request" attack (DoS attack).

## *2.3 Mathematical Modeling*

Network model: heterogeneous WSN, terrestrial nodes of which are randomly distributed in a certain area, have the same functions and resources, stationary, not serviced, do not change their location, equipped with a positioning system (GPS or another).

Assume that FICR and sensor nodes are equipped with the same radio equipment and support the same information exchange protocols (for example, IEEE 802.11), have limited radio communication range and throughput. Each network node has its own control system, operates in cooperation with FICR (if necessary, with other network nodes), has a sufficient amount of memory to store monitoring data.

FICR has the ability to move in three dimensions with constant or variable speed at a limited height and limited time. FICR has its own control system, which allows you to make your own decisions in the absence of connectivity with the control center of network. When FICR enters the radio communication zone of the node, it sends it the collected monitoring data according to the accepted exchange model.

It is assumed that FICR has information about the coordinates of the nodes, which can be obtained in one of the following ways:

1. At the stage of deployment terrestrial networks with a determined location of nodes, the coordinates of each node at its location are fixed.

2. In case of accidental deployment nodes of the network, FICR performs the initial overflight of the territory covered by the ground nodes of the WSN, and collects coordinates data of the nodes under the assumption of the presence positioning system in the nodes. In this case, the flight route is built to cover the entire observation area. FICR collects both monitoring information and information on the status of nodes and clusters for further planning of management tasks during the flight.
3. In the presence of a connected topology of WSN nodes with a ground gateway, it is possible for the control center to collect information about the status and coordinates of the location of sensor nodes.

Assume that the FICR radio communication zone with nodes is a circle (cluster) of radius R, and the reading of information from WSN nodes can be performed at arbitrary points of the cluster.

It is necessary: to find the minimum quantity of points of information collection points (cluster centers) from the nodes of the WSN and the coordinates of these points in space.

The task is to cover the original set of points (nodes) of the network with a minimum quantity of circles of radius *R*. Given:

- coverage area **A**, set of sensors $S = \{s_1, s_2, \ldots, s_i, \ldots, s_n\}$, $i = 1 \ldots n$ and their location on the ground $X = \{x_1, \ldots, x_i, \ldots, x_n\}$;
- $h$—flight altitude FICR;
- $d_{\max}$—the range of radio communication between the sensor node and IAFICR under the assumption of the boundary model of the radio channel;
- R—the radius of the coverage area FICR, $R = \sqrt{d_{\max}^2 - h^2}$;
- set of FICR $U = \{u_1, u_2, \ldots, u_k, \ldots, u_K\}$, $k = 1 \ldots K$, projection of their location on the ground $Y = \{y_1, y_2, \ldots, y_k, .., y_K\}$.

It is necessary: to find the minimum quantity of FICR (clusters $C = \{C_1, C_2, \ldots, C_j, \ldots, C_J\}$, $j = \overline{1, J}$) and their location in space (centers of FICR coverage areas with radius R), to cover all network nodes (Fig. 2).

Mathematical problem statement of minimizing the quantity of clusters:

$$K \to \min, \tag{7}$$

when the restrictions are met:

$$\min_{y_1, y_2, \ldots, y_K} \max_{j} \min_{x \in C_j} |x - y_k| \le d_{\max}, k = \overline{1, K}, j = \overline{1, J}, \tag{8}$$

$$q_i \le q_{\max}, \tag{9}$$

$$K \le K_{\max}, \quad R = [R_{\min} \ldots R_{\max}], \; h = [h_{\min} \ldots h_{\max}], \tag{10}$$

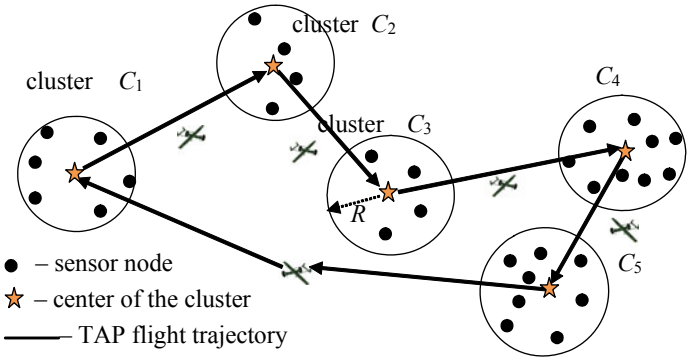



**Fig. 2** Option of network clustering and moving FICR between collection points—cluster centers

where $|x - y|$—Euclidean distance between points $x$ and $y$ on the ground, $q_j$—the quantity of sensor nodes in the $j$-th cluster. The physical meaning of the restrictions is as follows: inequality (8)—the maximum distance between the center of the coverage area and the sensor nodes must be minimized; inequality (9) determines the maximum quantity of nodes in clusters; (10)—sets resource constraints.

This task belongs to the class of computational geometry and close to the problem of placing $p$-centers - finding the minimum quantity of circles of fixed radius and their position, which cover a given quantity of points. The task is *NP*-complete, to obtain an exact solution of the task for large networks is difficult, so need to use heuristic search methods to solve it. Proposed to use iterative algorithms for cluster analysis of the formal element (for-el) and $k$-means to obtain a solution.

## 3 Simulation Results

Consider the simulation results for the following basic input data. Homogeneous nodes of WSN are located randomly on a certain plane. The quantity of sensor nodes—$N = 400$. The quantity of nodes in the cluster—$n_k = 10, 20, 50$. The initial energy of the nodes—$e_0 = 0.1$ J. Radio range—$d_{\max} = 250$ m, the maximum flight altitude IAFICR—$h_{\max} = 250$ m, maximum flight speed—$v_{\max} = 10$ m/s, quantity of flight circles (rounds)—$NR_{set} = 700$. IEEE 802.11 g channel access protocol, node monitoring data size—100 kb.

Consider and compare the following methods of data collection by relevant classes:

1. There is a method of direct collection in centroid clustering.
2. Improved (proposed in the article) method of direct data collection from FICR nodes in the implementation of various strategies (rules of flight and data exchange in clusters):

- strategy No. 1—FICR data collection point only in the center of the cluster;
- strategy No. 2—collection information by FICR taking into account the rules of flight over "critical" nodes;
- strategy No. 3—collection information by FICR taking into account the rules of exchange with nodes closer to the trajectory of the cluster;
- strategy No. 4—collection information by FICR from nodes during cooperative work on creation of mini-clusters and construction of energy-efficient routes to nodes that are closer to the flight route over cluster.

The efficiency of the methods will be compared according to the criteria: collection time of monitoring data $T_{\text{coll}}$ and time of operation WSN—$T_f$ (in physical terms - this is the time interval from the beginning to end of WSN work when spent all battery power of the last node as a result of information exchange. $T_f = 100\%$—this means that there are no working nodes for data transmission, i.e., in the last node the balance of energy consumption $e_{\text{con}}$ becomes zero. The higher value of $T_f$, the fewer working nodes remain in the network). Modeling and calculations were performed in the computer mathematics system MATLAB.

In Fig. 3 show the dependences of data collection time on the method of collection (centroid and advanced data collection methods) for different quantity of nodes in the virtual cluster $n_k = 10, 20, 50$. We can observe the advantages of the improved method of direct data collection. Data collection time with this method is reduced by 10–15% by reducing the quantity of flight points (application of the cluster analysis algorithm for-el).

In Fig. 4 show the dependences of WSN operation time on the method of collection (centroid and direct data collection) at different quantity of nodes in the virtual cluster $n_k = 10, 20, 50$. The improved method of direct data collection allows to increase the time of network operation by 12–17% due to the rules which apply for reducing energy consumption by nodes in exchanging with TAP).

In Figs. 5, 6, 7 and 8 show the results of modeling the application of the proposed rules for reducing the expense of nodes in the implementation of an improved method of direct data collection FICR. In these figures, the horizontal axis is the quantity of iterations, which means the quantity of flight rounds FICR.

The decrease in the average energy consumption by the $e_{\text{con}}$ node with the increase in the flight rounds is due to the fact that there are fewer working nodes. Because to the smaller quantity of nodes, the less opportunity to spend a lot of energy due to separate nodes. Therefore, the value of the average energy consumed by all nodes in the cluster is insignificant.

In Fig. 5 show the values of $e_{\text{con}}$ and $T_f$ for the strategy No. 1 (TAP flies over the center of the cluster), in Fig. 6—for strategy No. 2 (TAP flies over the node that has the lowest energy level and then through the center of the cluster).

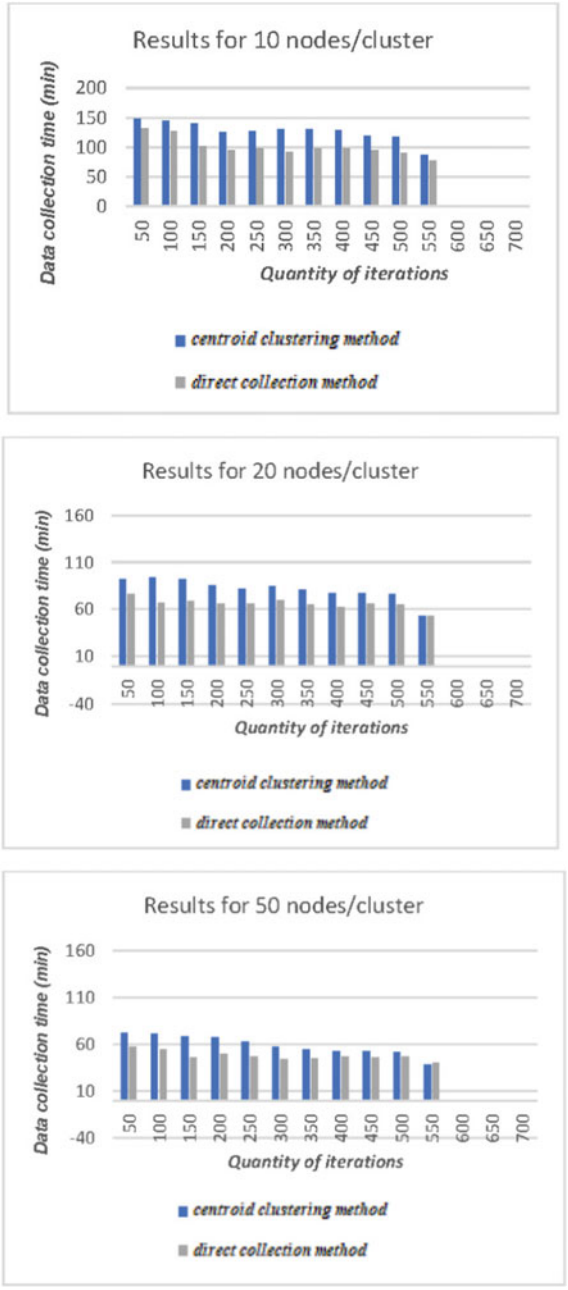



**Fig. 3** Dependence of data collection time on the method of collection with different quantity of nodes in the cluster $n_k = 10, 20, 50$




**Fig. 4** Dependence of WSN operation time on the collection method with different quantity of nodes in the cluster $n_k = 10, 20, 50$

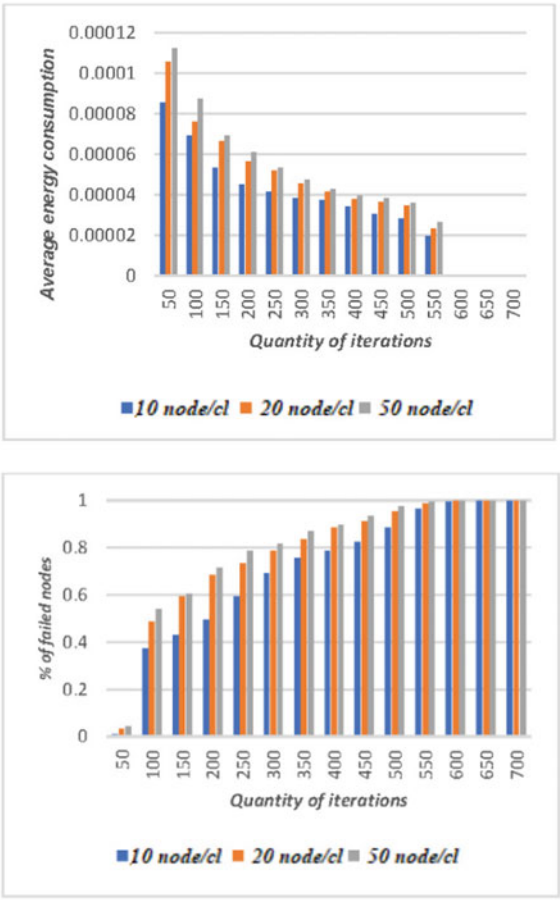



**Fig. 5** Average energy consumption of nodes and the percentage of failure nodes during the FICR flight over only cluster centers (strategy No. 1)

In Fig. 7 show the results of modeling the strategy No. 3 (nodes that are closer to the flight route of the cluster, transmit data).

From the simulation results, which are shown in Figs. 5, 6, 7, there is a general trend—with increasing rounds of data collection FICR, decreases the average energy consumption and increases the quantity of inoperable nodes in the network. This is explained by the following: in the initial rounds of flight, each cluster contains many working nodes, but this quantity decreases with the quantity of rounds of flight.

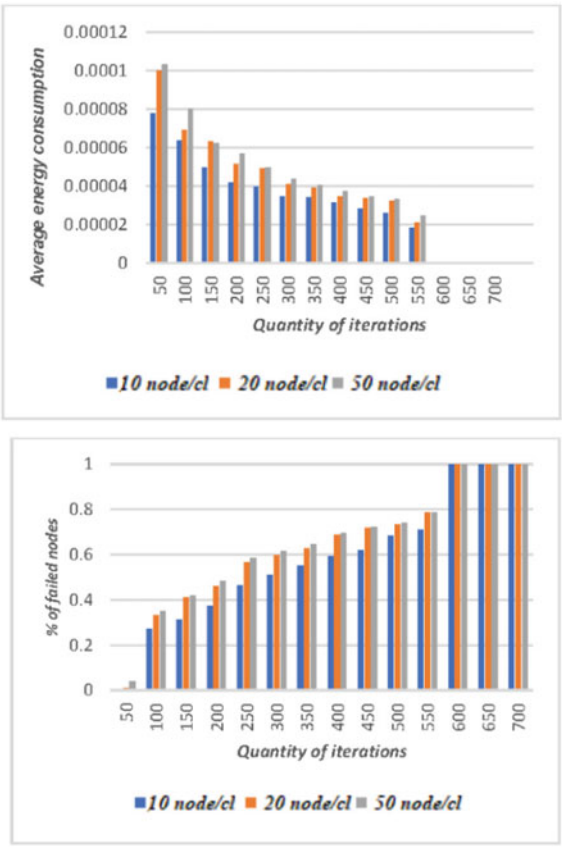



**Fig. 6** Average energy consumption of the node and the percentage of failure nodes during the flight of critical nodes (strategy No. 2)

The best results among the three strategies were shown by the strategy No. 3 (Fig. 5) (the lowest energy consumption and the lowest percentage of inoperable nodes) compared to the results in the previous two strategies.

The application of the cooperative strategy No. 4 (routing data from cluster nodes to the node closer to the FICR flight path using energy-efficient route search metrics) showed a significant advantage in low energy consumption and network operation time (Fig. 8).

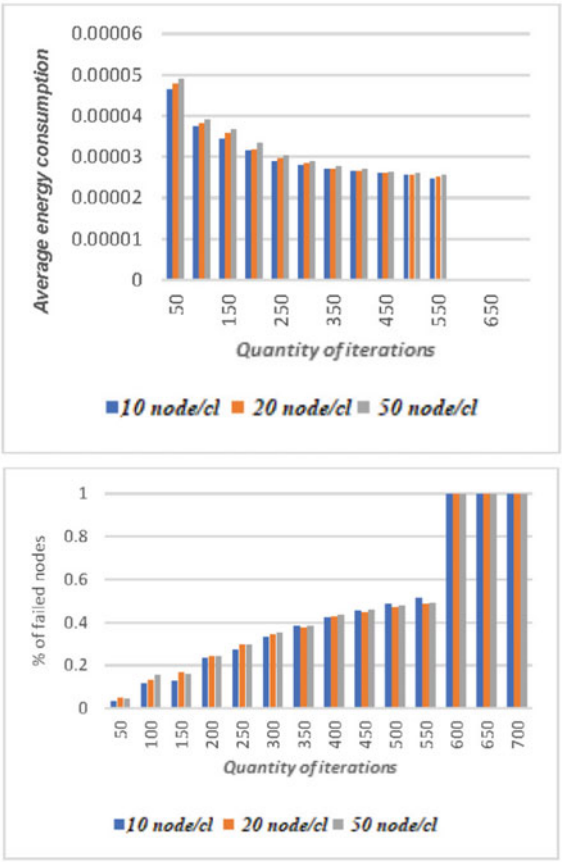



**Fig. 7** Average energy consumption per node and the percentage of failure nodes in a cooperative flight strategy (strategy No. 3)

Thus, after 600 rounds, in all three strategies the number of working nodes decreases sharply (strategies No. 1–3), but 40% of working nodes remain when using strategy No. 4. It is shown that a denser location of nodes will lead to less power consumption and longer operating time. This is the result of using energy-efficient route construction metrics to find a router with more battery power.

In Figs. 9 and 10 show the results of modeling the advanced method of data collection by FICR from the main nodes of the cluster in comparison with the method of its class (HEED) in indicators of data collection time and network operation time.

In the proposed method in comparison with the HEED method, the data collection time is reduced by an average of 14% due to the use and priority of the metric of the main node of the cluster—a shorter distance to the flight path FICR. The gain increases with the share of failed nodes. The energy consumption of nodes in real clustering in the proposed method is reduced by 10–15% due to the application

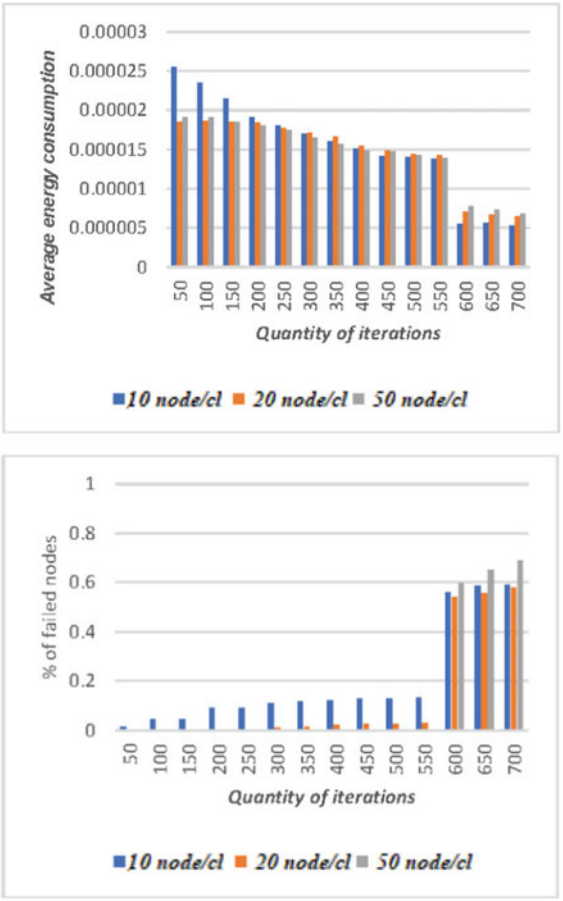



**Fig. 8** Average energy consumption and percentage of disabled nodes when using energy efficient metrics of route selection (strategy No. 4)

of energy-saving rules of cluster topology, equalization of energy consumption of nodes in the construction of transmission routes in the cluster (routes that have a minimum of energy consumption for transmission are selected from set of possible transmission routes between node and MNC and nodes whose battery level does not exceed the limit level).

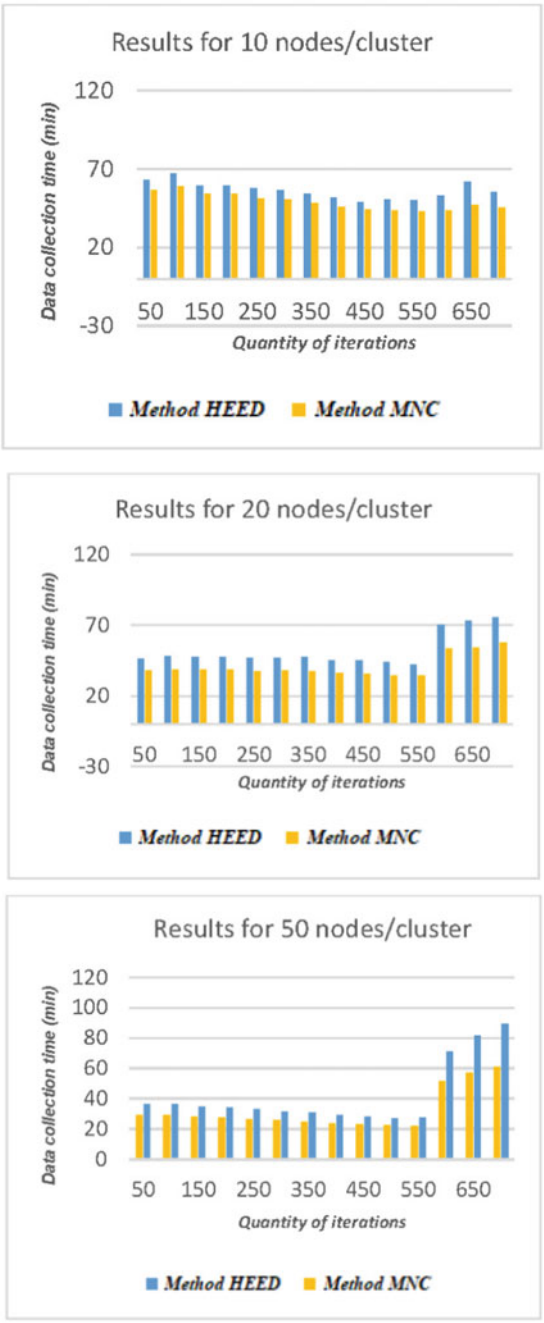



**Fig. 9** Dependences of data collection time from WSN by FICR' for HEED method and proposed improved method of collection from main nodes

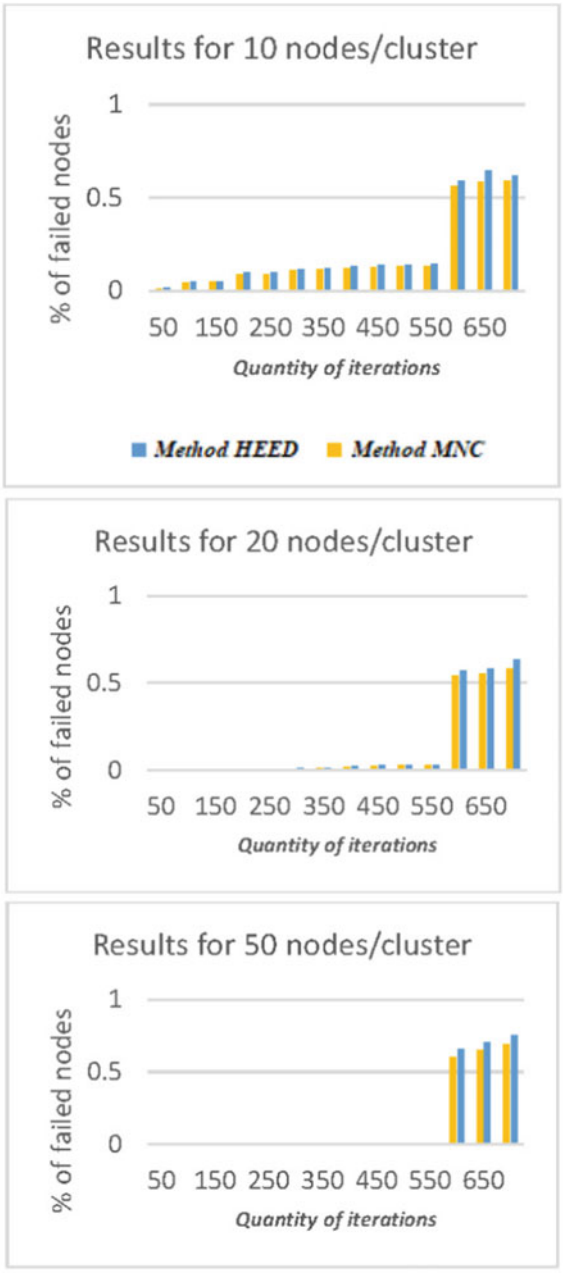



**Fig. 10** Dependencies of WSN operation time for HEED method and proposed method of collection from main TAP nodes

## 4 Conclusions

1. Conducted research of the effectiveness of the improved method of collection monitoring data with different source data: network size, quantity of clusters, quantity of nodes in the cluster, options for constructing data collection methods, flight strategy over nodes in the cluster.
2. The results of comparative modeling of the advanced direct method of data collection by FICR with the existing centroid method of data collection monitoring showed that the time of data collection from nodes by FICR using the improved method is reduced by 10–15% due to the reduction of flight points (application of cluster analysis algorithm for-el). The advanced method allows to increase the operation time of the network by 12–17% due to the application of the rules of reducing energy consumption by nodes when exchanging with FICR. The analysis of four cluster flight strategies (only between the centers of collection points; flight over critical nodes; transmission at points closer to the flight route; cooperative) showed the advantages of a cooperative strategy, which allows to reduce energy consumption of nodes in clusters by up to 15%.
3. Conducted evaluation of the efficiency of the improved method of data collection from the main nodes of the clustered network showed its advantages over the existing methods of this class. In comparison with the HEED clustering method, the improved method allows to reduce the time of monitoring data collection by an average of 14%, to increase the network operation time by 10–15%.

## References

1. Popescu D, Stoican FL, Stamatescu G, Chenaru O, Ichim L (2019) A survey of collaborative UAV–WSN systems for efficient monitoring. Sensors 19(21):4690. https://doi.org/10.3390/s19214690
2. Bunin SG, Woiter AP, Ilchenko MU, Romanyuk AV (2013) Self-organizing radio networks with ultra-wideband signals, Scientific Thought of the NAS of Ukraine, p 444. NPP Publishing House, Kyiv
3. Romanchenko IS, Danilyuk SL, Chumachenko SM et al (2016) Models of application of information and telecommunication technologies on the basis of unmanned aviation complexes in emergency situations, p 232. In: NAU, Kyiv
4. Romanyuk AV (2018) Algorithm for temporary clustering of nodes of wireless sensor networks for collecting monitoring information using UAVs, no 2 (33), pp 106–117. Interdepartmental scientific and technical collection Adaptive automatic control systems. https://doi.org/10.20535/1560-8956.33.2018.164680
5. Romanyuk AV (2018) Method of access to the radio channel by nodes of the wireless sensor network when collecting monitoring data by telecommunication air platforms, no 4, pp 84–91. Collection of scientific works of VITI
6. Romanyuk AV (2018) Tasks of UAV monitoring data collection management in wireless sensor networks, no 2, pp 103–112. Collection of scientific works of VITI

7. Romanyuk AV (2017) Target functions of nodes control of wireless sensor networks for monitoring critical objects of infrastructure, Scientific notes of TNU named after Vernadsky V.I., vol 28 (67), no 2, pp 49–54. Series: Technical Sciences, ISSN 1606-3721
8. Lysenko O, Valuiskyi S, Kirchu P, Romaniuk A (2013) Optimal control of telecommunication airplatforms in the area of emergency. Telecommun Sci 4(1):14–20

# Wireless Connection of Drones to the Base Station of the Existing Terrestrial Mobile Network

**Serhii Kravchuk and Irina Kravchuk**

**Abstract** A solution to the problem of finding the location and orientation of the repeater drone antenna system in the area of the radiation pattern of the base station (BS) antenna system by developing a new three-dimensional model of such a scenario is proposed; Based on the model, an analysis of the wave propagation losses over wireless channels between a ground BS, a repeater drone and a user terminal is carried out. A new approximation of the BS radiation pattern in the horizontal and vertical planes, the axes of which are interconnected by the geometry of the ellipse, has been developed. The plane of such an ellipse corresponds to the cross-sectional plane of the BS directional pattern with the position of the drone, which forms a radio channel with the BS and the user terminal. The results of modeling the scenario BS-drone-terminal are presented in the form of dependences of path losses of Line of Sight (LoS) and Non Line of Sight (NLoS) on the distances between the ground terminal, the BS, and the air platform at various values of hovering of the latter.

**Keywords** Unmanned aerial vehicles · Drone · Base station antenna · Communication system · Terrestrial mobile network · Wireless technologies

## 1 Introduction

The development of cellular communication systems is aimed at increasing their throughput and intellectualization, as well as increasing the possible number of different types of serviced mobile terminals. The implementation of the fifth and sixth generations of mobile technologies (5G and 6G) should connect people, things, data, applications, transportation systems, and cities into smart information and communication infrastructure [13]. At the same time, huge amounts of data should be transferred much faster, reliable connections to a huge number of devices should be provided, and large amounts of data processed with minimal latency.

S. Kravchuk (✉) · I. Kravchuk
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Ave. 37, Kyiv 03056, Ukraine
e-mail: sakravchuk@ukr.net

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_19

Currently, the need for a permanent or temporary expansion of the coverage area of a communication network can be realized not only through satellites or mobile base stations (BS) but also through more affordable solutions, for example, unmanned aerial vehicles (UAVs) or drones [4–6]. Drones can be part of the telecommunications infrastructure of telecom operators, performing the function of relaying telecommunications signals. The feasibility of using drones to expand the coverage of telecommunications services is associated with the relatively lower cost of such solutions in comparison with traditional base stations, the absence of the need for investment in-ground infrastructure, and greater versatility of possible applications [7].

Drones can also be used when planning a cellular network. Drone data allows you to accurately determine the optimal place where a new base station should appear—so that it covers a larger area and gives a better signal. Information from the drone helps to choose the optimal configuration for the base station—to determine at what height the antennas should be mounted, how many sectors are required, what type of antennas are needed, what should be the antenna tilt, and azimuths.

A new term Aerial base station (ABS), also known as an UAV-mounted BS [8], has even appeared in technical use. Because the coverage area of ground base stations is limited by various obstacles in the form of city buildings and terrain, a base station based on a drone has a much greater likelihood of maintaining communication with ground terminals. In this case, the likelihood of the presence of line-of-sight (LOS) conditions with them will be high. In addition, the installation of base stations on the drone without degrading the throughput of a given service area allows you to reduce the number of base stations of the ground mobile communication system.

The use of drones to form a separate cell in a mobile network involves two scenarios: a drone as a base station and a drone as a repeater between the base station and the user terminal (see Fig. 1).

In the first scenario, the coverage area of the cell is determined by the antenna system located on the flying drone. In this case, telecommunications equipment should be located onboard the drone, which will support the functioning of a full-fledged BS and its wireless connection both with the terrestrial network and with users. At the same time, wireless channels use different frequencies and technologies to communicate with the core network and to form the coverage area. However, such "versatility" requires an increase in the energy and overall performance of

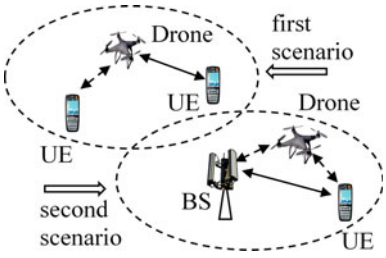



**Fig. 1** Scenarios for using drones to form a cell in a mobile network

the onboard equipment and, thus, significantly increases the cost of the drone-BS communication system. In addition, in this scenario, there is a traditional centralized point-to-multipoint scheme of access to user terminals, which cannot deal with shadow spots in its service area.

In the second scenario, which will be the focus of this work, the drone performs, as it were, additional functions for the operation of the ground BS. These functions can be expressed in improving the coverage area of the cell, increasing the likelihood of line-of-sight for the operation of user terminals, reducing the shading area, creating an opportunity to use technologies to increase the energy of radio channels, for example, cooperative relaying. Here, the requirements for onboard equipment of the drone in terms of mass-energy and dimensions are significantly lower than for equipment in the first scenario. The drone can perform such auxiliary functions not only for one BS but also for several neighboring BSs of the same terrestrial network. The mobility of the drone and its independence in the second scenario makes it possible to implement: a dynamic three-dimensional system, where the repeater on the drone can have a direct relationship with many base stations and terminals, line of sight, which is almost independent of the building and terrain, simple reconfiguration to other base stations. However, to implement this scenario, there is a problem of finding the exact location and orientation of the antenna system of such a drone in the radiation area of the BS antenna system.

The purpose of this work is to resolve the presented problem of finding the location and orientation of the repeater drone antenna system in the area of the BS antenna system by developing a new three-dimensional model of such a scenario; on the basis of such a model, analysis of path losses over wireless channels between a ground BS, a repeater drone and a user terminal.

## 2 Related Works

At present, the direction considered in the work is strongly attracting the attention of many researchers. For example, [9] provides an overview of UAV communication with 5G/6G wireless networks. UAVs were classified according to different areas and levels, including physical and network levels. It is noted that compared to communications with fixed infrastructure, UAVs add such advantages as flexible deployment, reliable LOS lines, and additional structural degrees of freedom with controlled mobility. UAVs are expected to become an important component of future wireless networks that can potentially support high-speed data transmission.

In [10], the concept of three-dimensional (3D) cellular networks is presented, which unite BS based on drones (drone-BS) and drone users (drone-UE). For this three-dimensional cellular architecture, a network planning approach is proposed that proposes a guided drone-BS deployment method based on the concept of truncated octahedron shapes. This method provides complete coverage for a given space with a minimum number of drone base units. To characterize frequency planning in such three-dimensional wireless networks, an analytical expression for possible frequency

reuse factors has been proposed. Simulation results show that the proposed approach reduces the latency of unmanned UEs compared to the classical cell association approach, which uses a signal-to-interference-plus-noise ratio (SINR) criterion. In particular, the proposed approach results in an average latency reduction of up to 46% compared to the SINR-based association. However, this work does not pay attention to the peculiarities of the drone entering into communication with the base station of the network.

In [11], the applicability of the millimeter-wave range for the UAV-BS system in a densely crowded area was evaluated using the NS-3 modeling tool. Simulation of data transmission with a transmission rate of 100 Mbps at a frequency of 28 GHz between a BS and a user terminal (UT) has been carried out. The BS and UT are located in opposite corners, and the UAV is in the middle of the service area. Retransmission using UAV-BS showed a performance gain of up to 80% for a scenario with high blocking. However, here, in the calculations, the probability of the UAV position falling into the coverage area of the base station antenna was not taken into account.

In [12], a statistical model of wave propagation is proposed to predict losses on the air-ground path between a low-altitude air platform and a ground terminal. The forecast is based on taking into account the properties of the urban environment and depends on the elevation angle between the terminal and the platform. However, the model is "flat" in nature and does not take into account all possible hovering positions of the UAV.

In [13], several modern Pathloss models were analyzed, applicable in scenarios that take into account the influence of the indoor (in buildings) and outdoor environments. The simulation results show that the standard Winner II model with an optional indoor blocking component provides the best path loss model for an urban emergency scenario. In addition, the influence of various topological parameters, such as the hovering height of the UAV, the height of the building, the internal and external distances of users, on the characteristics of the path loss was also analyzed. However, this analysis is rather approximate.

As a technique for providing wireless services to terrestrial users for various scenarios, [14] proposes the use of DSC small cells (drone small cells), which are served by airborne wireless base stations installed on a drone. The optimal hover heights of the drone have been determined, which leads to the maximum coverage on the ground surface and the minimum required transmit power for one DSC.

In [15], the optimal sizes were determined for small cells formed based on a drone. The traditional loss models (improved model) have been extended to include transmitter antenna gain patterns and wireless multipath fading. It is shown that for the improved model, there is an optimal hover height of the drone that provides maximum performance, taking into account the likelihood of failure, bit error rate, and throughput.

[16] analyzes the efficient deployment of multiple UAVs with directional antennas that act as wireless BSs to provide coverage on the Earth's surface. First, the probability of the downlink coverage for the UAV is determined as a function of the antenna height and gain. The three-dimensional locations of the UAV are then determined to maximize the overall coverage while maximizing the UAV's lifespan. The minimum

number of UAVs required to guarantee the target coverage probability for a given geographic area has been determined.

In [17], the architecture of radio access networks based on a drone is proposed, in which the latter is used to transfer data between BS and users. Based on current models of radio channels between drone-user and drone-based station (D2B), the service area of the users is analyzed. It then evaluates the ability to maximize user coverage while maintaining D2B communication quality for a given number of drone cells being deployed. It uses the Particle Swarm Optimization (PSO) algorithm.

The goal of [18] was to develop a prototype control system of the LTE (Long Term Evolution) standard for low-altitude drones and to study the possibility of implementing cellular communication based on drones. Parameters such as latency, handover, and signal strength were measured at different hovering altitudes. The measurement results showed that with an increase in the height of the drone hovering from the ground to 170 m, the received signal power decreased by 20 dBm, the downlink data rate decreased to 70%, and the delay increased to 94 ms. Thus, while the existing LTE network can provide the minimum required to support drone-based cellular communications, further research to increase aerial coverage, eliminate interference and reduce network latency remains relevant.

In [19], the deployment of a drone base station is considered without a priori information about the distribution of users. A sweep and search algorithm is proposed for finding the optimal deployment of drone base stations. The algorithm includes polygon area decomposition, coverage control, and collision avoidance.

In [20], the results of experimental testing and development of the basic principles of control of a swarm of drones and the formation of a cooperative relay system on their basis are presented. Scenarios of centralized and distributed construction of a collective control network for a swarm of drones for communication services have been worked out. The drone swarm was tested according to two relay scenarios: passive using metalized reflectors and active using additional radio units.

In [21], the case of using a UAV as a mobile Machine-to-machine (M2M) gateway for wireless all-penetrating sensor networks within a distributed control system is considered. The structure of such a pervasive distribution network is presented.

In [22], the possibilities of using radio frequency identification technology when a swarm of drones operates in a communication system mode are presented, that is, the use of RFID (Radio Frequency Identification) for the formation of control channels and traffic of a distributed self-organizing system of drones. The procedure for the exchange of control commands in a swarm using RFID is described.

In [23–25], the principles of the functional and structural design of the communication part of the UAV system based on SDR (Software-Defined Radio) and SoC (System-on-a-chip) technologies are presented with ensuring the survivability of the generated radio channels. Some new wireless communication technologies are applicable to improve the formation of radio channels in an integrated UAV-based station-user terminal system. Descriptions of such techniques as cooperative relaying, passive reflection are considered in [26–30].

The analysis of the results obtained by scientists and engineers in the research and development of mobile communication systems using a repeater drone is still

relevant. Such research will make it possible to obtain innovative technical solutions, improve existing communication technologies, create new software products, mathematical models, methods, algorithms.

## 3 The Scenario of the Interaction of a Drone with a Ground Mobile Network

In the scenario under consideration, there is a wireless interaction of the BS of the existing terrestrial mobile communication network with a flying repeater drone, the task of which is to improve the coverage in the cellular area of the BS operation. The drone can have onboard radio equipment of two types: an active receive-decode-transmit system and a passive system without decoding, mainly based on a reflective antenna system. The objects of such interaction through wireless communication channels are the antenna systems of BS (BS1 and BS2), drones (UAV1 and UAV2), and user terminal UE1 (Fig. 2). A necessary condition for the implementation of the interaction between them may be the mandatory location of antenna systems in the radiation zones of the directional patterns of each of their objects. If ground objects (BS and terminal) can be fixed in their locations, then the drone is mobile in all spatial coordinates and, thus, determines the possibility of interaction between all objects. Therefore, an important task is to determine the hover height and tilt of the drone's antenna plane (position) so that it can provide a signal to the user from the BS. Moreover, here not only direct signal retransmission is possible, but the implementation of cooperative retransmission.

A flat two-dimensional geometry of the scenario for connecting a drone to a cellular communication system to support the radio channel of a base station with a user terminal is shown in Fig. 3. Despite the simplicity of this geometry, the latter makes it possible to visually determine important parameters for further scenario modeling. These include: $\alpha_{bs}$ ($\psi_z$) and $\alpha_{dr}$—the angles of deviation from the horizontal line of the BS directional pattern and the plane of the reflective antenna of the drone, respectively; $2\theta_{bs}$ is the opening angle of the BS antenna at the level of −3 dB;

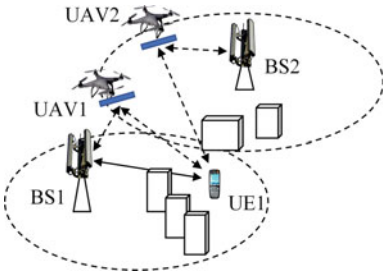



**Fig. 2** The scenario of interaction between ground base stations BS1 and BS2, repeater drones UAV1 and UAV2, and user terminal UE1

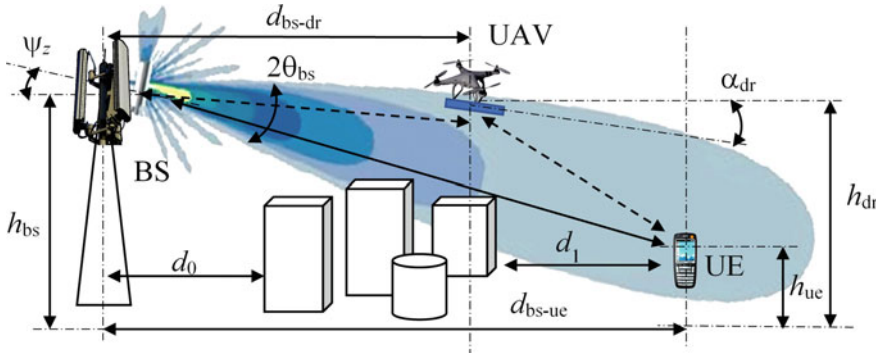



**Fig. 3** Two-dimensional representation of the UAV connection to the cellular communication system to support the BS radio channel with the user terminal UE

$d_{ue}$, $d_{dr}$, $d_0$—distances from the BS to the first obstacle, drone and user terminal, respectively; $d_1$ is the distance from the user terminal to the first obstacle; $h_{bs}$, $h_{dr}$ and $h_{ue}$—heights of the position of the suspension point of the antenna of the BS, drone and user terminal, respectively.

Many different models have now been developed to estimate the propagation loss of an air-to-ground radio signal. All of these models are designed to solve certain narrow tasks, the solution of which allows you to implement completely different scenarios for the operation of radio equipment on the air-ground route (service radio lines for avionics control, monitoring, navigation, radio data transmission lines, inter-platform communication, etc.) [31, 32]. This paper solves the problem of analytical selection of a fairly simple model for calculating the signal propagation loss to support radio links from a ground terminal to air platforms (heights up to 1000 m) and its analysis.

The proposed approach to modeling the air-to-ground propagation channel is to separately consider the line of sight and non-NLOS components together, taking into account their potential for occurrence. Note that in the case of NLOS, due to the effect of shadowing and reflections of signals from obstacles, the loss on the path is higher than in the case of LOS. Therefore, in addition to the free space LOS propagation loss, additional loss values are assigned to the NLOS radio link (Fig. 4). Thus, in general terms, the average losses on the air-ground radio path can be written as the following function: $\gamma_{av} = f(\gamma_0, p_{LOS}, \xi)$, where $\gamma_0$ is the propagation loss in free space; $p_{LOS}$ is the probability of the presence of a line-of-sight radio path at an elevation angle $\theta$; $\xi$—average additional loss (to loss in free space) during propagation, which depends on the environment.

For free space, the propagation loss is known to be $\gamma_0 = 20 \lg (4\pi f_c d/c_0)$, dB, where $f_c$ is the carrier frequency, Hz; $c_0$ is the speed of light; $d = (R^2 + h^2)^{1/2}$ is the distance between the drone and the ground terminal; $R$ is the ground distance between the drone and the ground terminal; $h$ is the hovering height of the drone.

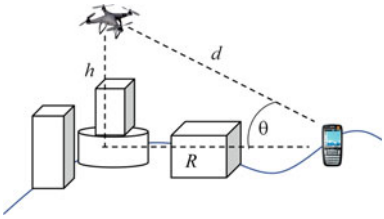



**Fig. 4** The scenario of the radio link drone-ground terminal ("air-ground")

The probability of a line-of-sight radio path according to [33] can be determined from the following equation $p_{LOS} = 1/(1 + \alpha \times \exp(-\beta \times (180/\pi\,\theta - \alpha)))$, where $\alpha$ and $\beta$ are constant values that depend on the environment (rural, urban, dense buildings). For example, for a city, you can take the values $\alpha = 9.6$ and $\beta = 0.28$.

Then the loss for the propagation of radio waves in conditions of line of sight can be written as $\gamma_{LOS} = p_{LOS}\,(\gamma_0 + \xi_{LOS})$, and in conditions of indirect visibility - $\gamma_{NLOS} = (1\text{-}p_{LOS})(\gamma_0 + \xi_{NLOS})$. The values $\xi_{LOS}$ and $\xi_{NLOS}$ have quite different meanings. So $\xi_{LOS}$ accepts units of dB and mainly takes into account insignificant shadowing on the LOS path, while $\xi_{NLOS}$ is tens of dB and takes into account the presence of surrounding buildings, relief elements, and vegetation shadows. Based on the expressions given in [34, 35], we can write the following equation for $\xi_{NLOS} = 20 \log_{10}(2 \sin(2\pi h_t h_r/\lambda/d))$, where $\lambda$ is the wavelength; $h_t$ and $h_r$ are the heights of the transmitter and receiver, respectively; $d$ is the distance between transmitter and receiver. Based on the above, we can write the following expression for the propagation loss.

$$\gamma_{av} = f(\gamma_0, p_{LOS}, \xi) = \gamma_{LOS} + \gamma_{NLOS} \tag{1}$$

## 4 Three-Dimensional Model of the Formation of Radio Links Between the Antenna Systems of the BS and the User Terminal with Retransmission Through the Drone

Consider a three-dimensional coordinate system in which the position of the BS antenna system is in the center of the plane (x, y), and the position of the UE in this plane is measured only along the x-axis, that is, its y-coordinate is zero (Fig. 5). Here the plane (x, y) is the flat surface of the Earth.

Three points of the antenna systems of the BS, drone (UAV), and UE form an area in three-dimensional space with coordinates.

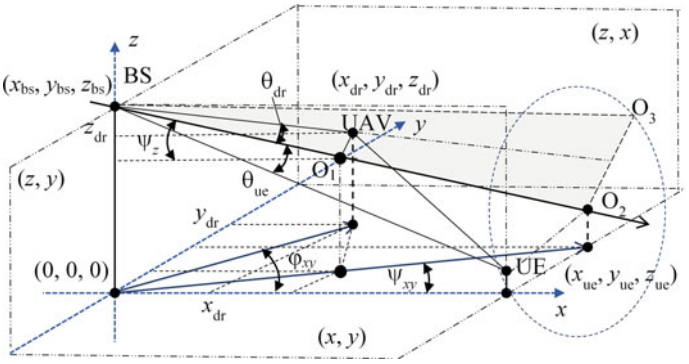



**Fig. 5** Three-dimensional model of radio links between antenna systems BS and UE with retransmission through the drone UAV

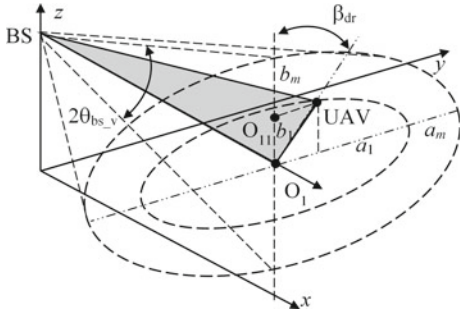



**Fig. 6** The scenario of intersection of the directional pattern of the BS antenna with the elliptical plane where the drone is located (UAV—the point of hovering of the drone)

$$\begin{Vmatrix} x_{\text{bs}} & y_{\text{bs}} & z_{\text{bs}} \\ x_{\text{ue}} & y_{\text{ue}} & z_{\text{ue}} \\ x_{\text{dr}} & y_{\text{dr}} & z_{\text{dr}} \end{Vmatrix},$$

where $x_i$, $y_i$ and $z_i$ are three-dimensional coordinates BS ($i = \text{bs}$), UE ($i = \text{ue}$) and drone ($i = \text{dr}$), respectively.

We introduce the following designations for our three-dimensional scenario: $\psi_{xy}$ is the angle between the $x$-axis and the projection onto the plane $(x, y)$ of the vector of the main axis of the antenna pattern of the BS antenna; $\psi_z$ is the angle between the plane, which is parallel to the plane $(x, y)$ and also passes through the suspension point of the BS antenna at the height $h_{\text{bs}}$, and the vector of the main axis of the BS antenna directional pattern; $\varphi_{xy}$ is the angle between the $x$-axis and the projection onto the plane $(x, y)$ of the vector from the point of the BS antenna in the direction to the drone; $h_{\text{bs}}$, $h_{\text{ue}}$ and $h_{\text{dr}}$—the height of the BS, UE and drone antenna systems,

respectively; $l_{ue}$ and $l_{dr}$ are the distances along the ($x$, $y$) plane from the BS to the UE and the drone, respectively.

Then, from the three-dimensional geometry of the scenario, we determine that the BS coordinates will be ($x_{bs} = y_{bs} = 0.0$; $z_{bs} = h_{bs}$), the UE coordinates will be ($x_{ue} = l_{ue}$, $y_{ue} = 0.0$, $z_{ue} = h_{ue}$) and the drone coordinates will be ($x_{dr} = l_{dr} \cos \varphi_{xy}$, $y_{dr} = l_{dr} \sin \varphi_{xy}$, $z_{dr} = h_{dr}$).

The slant range will be between:

- BS and drone:

$$d_{bs-dr} = [(x_{dr}-x_{bs})^2 + (y_{dr}-y_{bs})^2 + (z_{dr}-z_{bs})^2]^{1/2} = \left[x_{dr}^2 + y_{dr}^2 + (z_{dr}-z_{bs})^2\right]^{1/2},$$

- BS and UE:

$$d_{bs-ue} = [(x_{ue}-x_{bs})^2 + (y_{ue}-y_{bs})^2 + (z_{ue}-z_{bs})^2]^{1/2} = [x_{ue}^2 + (z_{ue}-z_{bs})^2]^{1/2},$$

- drone (UAV) and UE:

$$d_{dr-ue} = \left[(x_{dr}-x_{ue})^2 + (y_{dr}-y_{ue})^2 + (z_{dr}-z_{ue})^2\right]^{1/2} = \left[(x_{dr}-x_{ue})^2 + y_{dr}^2 + (z_{dr}-z_{ue})^2\right]^{1/2}.$$

Next, you need to assess the position of the drone and the UE in relation to the volume of the BS antenna pattern. This is important since the implementation of our scenario is possible only if the drone gets into the irradiation of the BS antenna system. The area of the radiation pattern is represented in the form of a straight cone with the apex at the point of the BS antenna system, the opening angle $2\theta_{bs}$, which is equal to the width of its main beam at the level of 0.5 of its maximum power values, and the line of the cone height coinciding with the main axis of symmetry BS antenna directional patterns. The area of the triangle (BS, $O_2$, $O_3$), the rotation of which around the main axis of the radiation pattern forms the BS radiation cone, is the plane of the analysis of the position of the drone in the radiation beam of the BS antenna. Let us define the ordinates of point $O_1$ as the point of the smallest distance from the drone to the main axis of symmetry of the BS antenna. Proceeding from the fact that the BS antenna serves as the point of the radiation source, the distances to UAV and $O_1$ are taken equal (the same radius of the wave front from the BS), that is, $\rho = |BS, UAV| = |BS, O_1| = d_{bs\text{-}dr}$. Then the coordinates of the point $O_1$ ($x_{o1}$, $y_{o1}$, $z_{o1}$) can be defined as.

$$y_{o1} = \rho_x \sin \psi_{xy},$$

$$x_{o1} = \rho_x \cos \psi_{xy},$$

$$z_{o1} = z_{bs} - \rho \sin \psi_z,$$

where $\rho_x = \rho \cos \psi_z$ is the projection of the vector $\rho$ onto the plane ($x$, $y$).

The distance between UAV and $O_1$ will have the form.

$$d_{\text{dr}-\text{o1}} = [(x_{\text{o1}}-x_{\text{dr}})^2 + (y_{\text{o1}}-y_{\text{dr}})^2 + (z_{\text{o1}}-z_{\text{dr}})^2]^{1/2}$$
$$= [(x_{\text{o1}}-l_{\text{dr}}\cos\varphi_{xy})^2 + (y_{\text{o1}}-l_{\text{dr}}\sin\varphi_{xy})^2 + (z_{\text{o1}}-h_{\text{dr}})^2]^{1/2}.$$

From the analysis of the plane of an isosceles triangle (BS, UAV, $O_1$), the angular position of the drone in the plane (BS, $O_2$, $O_3$) is determined with respect to the segment | BS, $O_1$ | and it is equal to $\theta_{\text{dr}} = \arccos(\rho/2\ d_{\text{dr-o1}})$. Now, knowing the opening angle of the $2\theta_{\text{bs}}$ BS antenna, we can determine the necessary condition for the location of the drone antenna in the BS radiation area: $\theta_{\text{dr}} < \theta_{\text{bs}}$.

A similar analysis can be done for the UE. Consider the plane (BS, $O_2$, UE) and determine the coordinates of the point $O_2$ ($x_{\text{o2}}$, $y_{\text{o2}}$, $z_{\text{o2}}$):

$$y_{\text{o2}} = \rho_{\text{x2}}\sin\psi_{xy},$$

$$x_{\text{o2}} = \rho_{\text{x2}}\cos\psi_{xy},$$

$$z_{\text{o2}} = z_{\text{bs}} - d_{\text{bs}-\text{ue}}\sin\psi_z,$$

where $\rho_{\text{x2}} = d_{\text{bs-ue}}\cos\psi_z$ is the projection of the vector $d_{\text{bs-ue}}$ onto the plane ($x$, $y$).

The distance between UE and $O_2$ will have the form.

$$d_{\text{ue}-\text{o2}} = [(x_{\text{o2}}-x_{\text{ue}})^2 + (y_{\text{o2}}-y_{\text{ue}})^2 + (z_{\text{o2}}-z_{\text{ue}})^2]^{1/2}.$$

From the analysis of the plane of an isosceles triangle (BS, $O_2$, UE), the angular position of UE with respect to the segment | BS, $O_2$ | and it is equal to.

$$\theta_{\text{ue}} = \arccos(d_{\text{bs}-\text{ue}}/2d_{\text{ue}-\text{o2}}).$$

Then the necessary condition for the location of the UE antenna in the BS radiation area: $\theta_{\text{ue}} < \theta_{\text{bs}}$.

Let us define the elevation (elevation) angles of the UE on the BS and the drone as $\alpha_{\text{ue_bs}}$ and $\alpha_{\text{ue_dr}}$, respectively. The elevation angles will be:

$$\alpha_{\text{ue_bs}} = \arcsin(|h_{\text{bs}} - h_{\text{ue}}|/d_{\text{bs}-\text{ue}}) \text{ and } \alpha_{\text{ue_dr}} = \arcsin(|h_{\text{dr}} - h_{\text{ue}}|/d_{\text{dr}-\text{ue}}).$$

The radiation pattern of the BS antenna system is sectorial with different distributions of the electromagnetic field in the vertical and horizontal planes. Hence, in the horizontal plane, the opening of the main beam of the directional pattern, mainly, has 60°, and in the vertical plane—up to 10°. For simplicity of analysis of the scenario under consideration, we will use the existing approximations of antenna radiation patterns. So, in the vertical plane, we will use the approximation for a narrowly directed diagram according to [36, 37]:

$$G_{\mathrm{v}}(\varphi) = G_{\max}\sin^2(\pi\, n\, \varphi)/(n \sin^2(\pi\, \varphi)),$$

where $n$ is an integer defining the steepness of the decrease in the directivity of the main lobe of the directivity pattern; $\varphi$ is the angle of deviation from the electrical axis (main lobe) of the antenna; $G_{\max}$ is the maximum value of the gain of the main lobe of the antenna. To approximate the horizontal directivity of the antenna main lobe characteristic, we use the expression from [38] $G_{\mathrm{v}}(\varphi) = G_{\max}\ (\cos\ \varphi)^n$. Since this expression does not reproduce the side-lobe structure, a flat side-lobe representation (usually specified by the antenna manufacturer) can be used for modeling as follows [39]:

$$G_h(\varphi) = \max[G_{\max}(\cos\varphi)^n, G_0],$$

where $G_0$ is the level of the first side lobes of the radiation pattern.

We are faced with the task of determining the position of the drone in relation to the plane of the cross-section of the directional pattern in order to determine the gain at a given point of hovering of the drone. Taking the elliptical section of the directional diagram, we find that the major and minor axes of the section ellipse are determined by the angular values $2\theta_{\mathrm{bs_v}}$ and $2\theta_{\mathrm{bs_h}}$, respectively. Then you can define the position of the drone antenna as a point on the arc of the ellipse of the section. To do this, we find the angle $\beta_{\mathrm{dr}}$ of the deviation of the drone antenna position relative to the vertical plane of the radiation pattern in the considered elliptical section. We determine the coordinates of the point $\mathrm{O}_{11}$ ($x_{\mathrm{o1}}$, 0, $z_{\mathrm{dr}}$) as the projection of the point UAV onto the perpendicular to the plane ($x$, $y$), which passes through the point $\mathrm{O}_1$, and the distance between UAV and $\mathrm{O}_{11}$:

$$d_{\mathrm{dr-o11}} = [(x_{\mathrm{o11}}-x_{\mathrm{dr}})^2 + (y_{\mathrm{o11}}-y_{\mathrm{dr}})^2 + (z_{\mathrm{o11}}-z_{\mathrm{dr}})^2\Big]^{1/2} = \Big[(x_{\mathrm{o11}}-x_{\mathrm{dr}})^2 + (y_{\mathrm{dr}})^2]^{1/2},$$

and the distance between points $\mathrm{O}_1$ and $\mathrm{O}_{11}$ will be.

$$d_{\mathrm{o1-o11}} = [(x_{\mathrm{o11}}-x_{\mathrm{o1}})^2 + (y_{\mathrm{o11}}-y_{\mathrm{o1}})^2 + (z_{\mathrm{o11}}-z_{\mathrm{o1}})^2]^{1/2} = |z_{\mathrm{o11}}-z_{\mathrm{o1}}|.$$

Then the deflection angle can be determined from.

$$\beta_{\mathrm{dr}} = \arcsin(d_{\mathrm{dr-o11}} d_{\mathrm{dr-o1}}).$$

The geometry of the ellipse is determined by: the semi-major axis $a$, the line of which is parallel to the axis (0, $y$), and the semi-minor axis $b$, the line of which coincides with the perpendicular to the plane ($x$, $y$) and passes through the point $\mathrm{O}_1$. For an elliptical section of the directional pattern of the BS antenna at a distance $d_{\mathrm{bs\text{-}o1}} = [(x_{\mathrm{o1}})^2 + (y_{\mathrm{o1}})^2 + (z_{\mathrm{o1}} - z_{\mathrm{bs}})^2]^{1/2}$ and at aperture angles $2\theta_{\mathrm{bs_v}}$ and $2\theta_{\mathrm{bs_h}}$ we have the size of large and small semiaxes $2a_m = d_{\mathrm{bs\text{-}o1}}/(2\cos(4\ \theta_{\mathrm{bs_}h}))$ and $2b_m = d_{\mathrm{bs\text{-}o1}}/(2\cos(4\ \theta_{\mathrm{bs_}v}))$, respectively. Hence, the ratio between the semiaxes can be determined as.

$$k_m = a_m/b_m = d_{\text{bs}-\text{o1}}/(2\cos(4\theta_{\text{bs_}h}))/d_{\text{bs}-\text{o1}}/(2\cos(4\theta_{\text{bs_}v})) = \cos(4\theta_{\text{bs_}v})/\cos(4\theta_{\text{bs_}h}).$$

For all ellipses that can be formed with the center at point $O_1$ and inside the area of the ellipse with $a_m$ and $b_m$, they will have $k_m = \text{const}$. This allows you to determine the ellipse of the cross-section of the radiation pattern (with semiaxes $a_1$ and $b_1$), on which the UAV antenna point is located. Let us reduce the three-dimensional coordinates UAV to two-dimensional ($x_{\text{dre}}$, $y_{\text{dre}}$) on the plane, which is formed by ellipses with the origin at the point $O_1$. Here $x_{\text{dre}} = d_{\text{dr-o11}}$, $y_{\text{dre}} = d_{\text{o1-o11}}$. Then the semiaxes of the perimeter of the ellipse where UAV is located can be found as $a_1 = (y_{\text{dre}}{}^2 k_m{}^2 + x_{\text{dre}}{}^2)^{1/2}$, $b_1 = a_1/k_m$. Now you can determine the opening angles of the diagram for an elliptical section with semiaxes $a_1$ and $b_1$ using the relations for an isosceles triangle.

$$2\theta_{\text{1bs_v}} = \arccos(d_{\text{v}}/(4b_1)),$$

$$2\theta_{\text{1bs_v}} = \arccos(d_{\text{v}}/(4a_1)),$$

where $d_{\text{v}}{}^2 = (d_{\text{bs-o1}})^2 - 4\,b_1{}^2$; $d_{\text{h}}{}^2 = (d_{\text{bs-o1}})^2 - 4a_1{}^2$; $d_{\text{bs-o1}} = [(x_{\text{o1}} - x_{\text{bs}})^2 + (y_{\text{o1}} - y_{\text{bs}})^2 + (z_{\text{o1}} - z_{\text{bs}})^2]^{1/2}$. Having determined the required operating angles of the BS antenna aperture, it is possible to determine the values of the gain at the points that lie at the ends of the semi-axes of the ellipse passing through the drone antenna: $G_h(\theta_{\text{1bs_h}})$ and $G_{\text{v}}(\theta_{\text{1bs_v}})$. Then: $G_{\text{vy}} = G_{\text{v}}\, y_1/b_1$; $G_{\text{hx}} = G_{\text{x}}\, x_1/a_1$; $x_1 = d_{\text{dr-o1}} \sin(\beta_{\text{dr}})$; $y_1 = d_{\text{dr-o1}} \cos(\beta_{\text{dr}})$; $G_{\text{dr}} = (G_{\text{vy}}\, G_{\text{hx}})^{1/2}$.

The simulation results based on the proposed model for approximating the directional pattern of the BS antenna in the Matlab package are shown in Fig. 7.

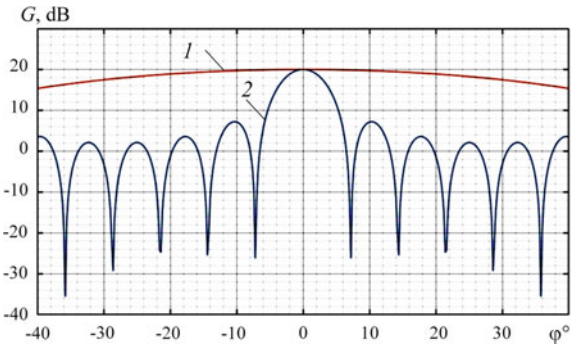



**Fig. 7** Approximation of the directional diagram of the BS antenna in the planes ($G_{\max} = 20$ dB): 1—horizontal $2\theta_{\text{bs_}h} = 60°$ ($n = 4$); 2—vertical $2\theta_{\text{bs_v}} = 6{,}4°$ ($n = 8$)

## 5 Simulation Results and Their Analysis

Scenarios were simulated on the Matlab software platform. For a two-dimensional scenario in Fig. 4, the results of modeling the path loss of radio waves from the distance between the ground terminal UE and the air platform UAV at various values of the hovering of the latter according to formula (1) are shown in Fig. 8. With an increase in the distance between UAV and UE, path losses increase, however, such an increase when taking into account losses only under the conditions of LOS (continuous curves in Fig. 8) and taking into account NLOS (dashed curves in Fig. 8) are noticeably different. In the case of path losses only under LOS conditions, a monotonic exponential increase in attenuation values is observed for all UAV hovering heights. Moreover, the higher the UAV is, the greater the path loss. However, it can be seen that at small ratios of the ground distance and hovering height, that is, high elevation angles θ, the signal propagation loss will be the lowest. In Fig. 8 also shows the dependences taking into account path losses under NLOS conditions, from which it can be seen that for an urbanized area, the higher the UAV is above the ground, the lower the value of path losses. This is due to an increase in the elevation angle and thus an increase in the LOS probability.

Qualitative characteristics of radio channels - signal-to-noise ratio SNR and bit error rate BER - for the scenario under consideration were calculated based on the following parameters: modulation type—QPSK; UAV transmitter power—100 mW; antenna gains UAV and UE, respectively—18 and 12 dB; carrier frequency—5.8 GHz; end-to-end bandwidth—100 MHz; receiver noise figure—6 dB. BER coefficients are calculated according to the < berawgn > function of the Matlab package. The simulation results are shown in Fig. 9. They demonstrate that maintaining the best radio link performance is possible with a higher drone position when the LOS

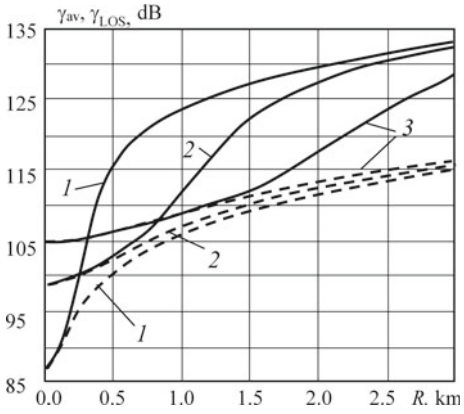



**Fig. 8** Dependence of pathlosses on the propagation of radio waves on the distance between the ground terminal and the air platforms at various values of hovering of the latter: *1*—100 m; *2*—400 m; *3*—800 m (continuous curves—taking into account losses in LOS and NLOS conditions, $\gamma_{av}$; dashed curves - taking into account losses only in conditions of LOS, $\gamma_{LOS}$)

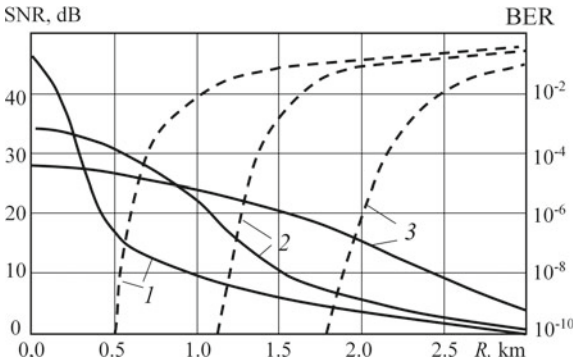



**Fig. 9** Dependences of the signal-to-noise ratio SNR (continuous curves) and BER (dashed curves) on the ground distance between UAV and UE at the UAV hovering height in m: *1*—100; *2*—400; *3*—800

component of the pathloss will dominate. So, for example, for a radio channel to have a BER of no worse than $10^{-6}$ at a ground distance of up to 1300 m, the UAV hover height is sufficient 400 m (curves *2* in Fig. 9), and up to 2000 m –800 m (curves *3* in Fig. 9).

For the three-dimensional scenario shown in Fig. 5, the following parameters were adopted: modulation type - QPSK; transmitter power at BS and UAV—100 and 50 mW, respectively; antenna gains UAV and UE - 10 and 12 dB, respectively; carrier frequency—5.8 GHz; end-to-end bandwidth—20 MHz; receiver noise figure—6 dB; the height of the BS and UE antennas is 15 and 1 m, respectively; the distance between the BS and the UE is 500 m. In this case, the position of the UAV is always between the BS and the UE, and the elevation angle $\alpha_{ue_bs}$ has a small value up to 2 degrees, which makes the UE characteristics poorly sensitive to small changes in the directional angle of the BS antenna; the directional pattern of the main beam of the BS antenna at a level of − 3 dB in the vertical and horizontal planes was 5 and 60 degrees, respectively; $\psi_{xy} = 0$ grad.

The appearance of a repeater on the UAV in the scenario immediately reduces the loss of signal transmission to the UE. Moreover, it can be seen that the location of the UAV allows you to selectively regulate the amount of propagation loss on radio links between it and the BS, or between the UAV and the UE. Thus, the curves numbered 2 in Fig. 10 demonstrate practically equal losses on radio links BS-UAV and UAV-UE, provided that the UAV is located in the middle between the BS and the UE (250 m). If the UAV hovering position is brought closer to the BS (150 m, curves in Fig. 10 numbered 1), the propagation loss on the BS-UAV radio link will decrease, and on the UAV-UE radio link, it will increase significantly. Everything will happen the other way around if the UAV hovering position is moved away from the BS (350 m, curves in Fig. 10 numbered 3). At UAV hovering heights above 15…20 m, nonlinearity of propagation loss characteristics are observed, caused by a different degree of probability of accounting for the total losses of LOS and NLOS components.

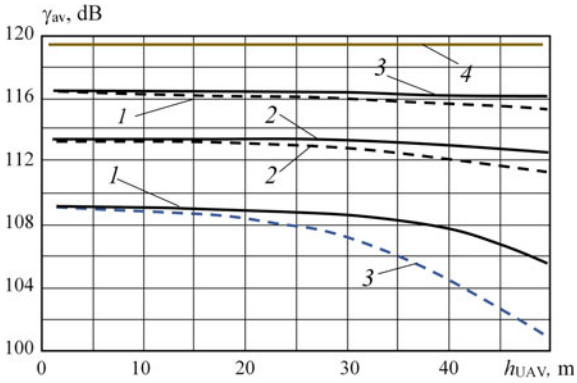



**Fig. 10** Dependence of the signal propagation pathloss on the UAV hovering height for the positions of the projection point of the UAV position on the earth's surface between the BS and UE (counting from the BS) 150 m (curves *1*), 250 m (curves *2*) and 350 m (curves *3*): solid curves correspond to pathlosses on the BS-UAV radio link; dashed curves—pathlosses on the UAV-UE radio link (curve *4*—total path losses on the BS-UE radio link)

Now let's move on to the analysis of the obtained electrical characteristics of the scenario under consideration, taking into account the influence of the UAV hovering position on the corresponding value of the gain of the BS antenna pattern (Fig. 7). In the proposed model, there is a functional relationship between the UAV hovering height and the angular coordinates $2\theta_{bs_v}$ and $2\theta_{bs_h}$. The simulation results in the form of the dependence of SNR at the UAV and UE receiver on the UAV hovering height for a number of ground distances between the BS and UAV are shown in Fig. 11. From the analysis of this figure, it can be seen that at the UAV hovering height corresponding to the BS antenna location ($h_{UAV} = 15$ m, the BS antenna directional pattern axis does not tilt to the earth's surface), there is a maximum SNR value at the UAV receiver (continuous and dotted curves Fig. 11), which corresponds to the maximum of the main lobe of the BS antenna, and the SNR characteristic itself near $h_{UAV} = 15$ m repeats the outline of the gain of the main lobe of the BS antenna. The closer the UAV is to the BS, the higher the SNR value at its receiver, and the lower the SNR at the UE receiver. For example, with $h_{UAV} = 15$ m and a distance from the BS to the UAV equal to 150 m, we have SNR = 22 dB for the UAV receiver, and SNR = 18 dB for a distance of 250 m. Similar behavior of the BER characteristics is observed on the UAV receiver (continuous curves in Fig. 12) and UE (dashed curves in Fig. 12) depending on the UAV hovering height.

At the heights of UAV operation above 20 m and below 30 m, there is a decline in the SNR characteristic for UAV, which corresponds to the "zeros" of the BS antenna pattern. A further increase in SNR values (Fig. 11) and a decrease in BER values (Fig. 12) correspond to the UAV rise heights, where its antenna system falls on the side lobe of the BS antenna pattern. All dependencies show that the further and higher the UAV is from the BS, the better the characteristics associated with the UE.

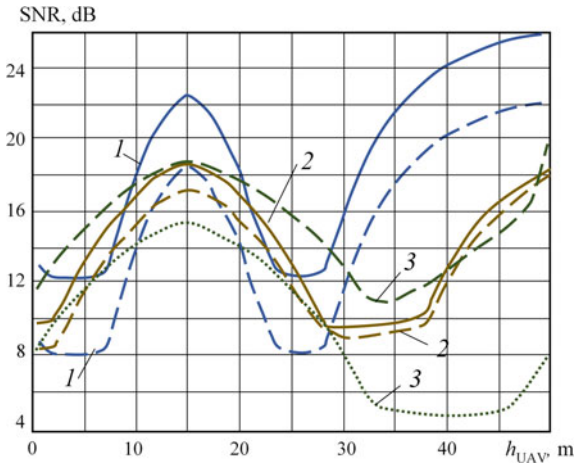



**Fig. 11** Dependence of SNR at the receiver of UAV (continuous and dotted curves) and UE (dashed curves) on the UAV hovering altitude for the following ground distances between the BS and UAV: *1*—150 m; *2*—250 m (middle of the radio link); *3*—350 m

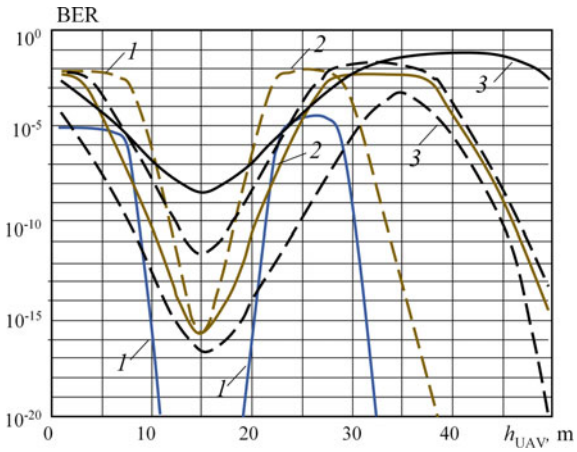



**Fig. 12** Dependence of BER at the receiver UAV (continuous curves) and UE (dashed curves) on the UAV hovering altitude for the following ground distances between the BS and UAV: *1*—150 m; *2*—250 m (middle of the radio link); *3*—350 m

The influence on the characteristics of the radio link that the UAV drifts away from the axis of the main lobe of the BS antenna pattern by the angle $\varphi_{xy}$ is shown in Fig. 13. Here, an increase in the deflection angle $\varphi_{xy}$ leads to a deterioration in the signal-to-noise ratio at the UAV input.

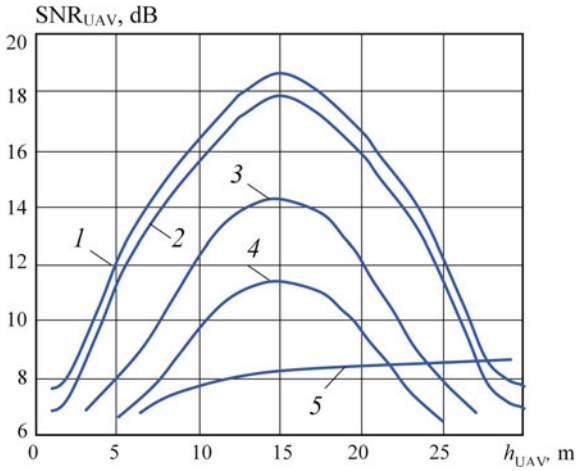



**Fig. 13** Dependence of the signal-to-noise ratio at the UAV input on the hovering height of the latter at the angle between the x-axis and the projection on the plane (x, y) of the vector from the point of the BS antenna in the direction to UAV $\varphi_{xy}$, deg: 1–0; 2–2.5; 3–3.75; 4–5; 5–7

## 6 Conclusions

A solution to the problem of finding the location and orientation of the antenna system of a repeater drone (or UAV) in the area of the radiation pattern of the BS antenna system by developing a new three-dimensional model of such a scenario is proposed; Based on this model, an analysis of the wave pathlosses over wireless channels between a ground BS, a repeater drone and a user terminal is carried out. A new approximation of the directional diagram of the BS in the horizontal and vertical planes, the axes of which are interconnected by the geometry of the ellipse, has been developed. The plane of such an ellipse corresponds to the cross-sectional plane of the BS directional pattern with the position of the drone, which forms a radio channel with the BS and the user terminal. The results of modeling the scenario BS-drone-terminal are presented in the form of dependences of pathlosses on the propagation of LOS and NLOS conditions on the distances between the ground terminal, the BS, and the air platform at various values of the hovering of the latter.

It has been determined that the best results for maintaining the performance of radio links in urban environments will be provided for UAV hovering heights equal to and above the main lobe axis of the BS antenna. Moreover, at high altitudes of UAV operation (outside the angular coordinates of the main lobe of the BS antenna directional pattern), the latter ensures the operation of the radio link using the upper side lobes of the BS antenna directional pattern. This is also facilitated by the presence of low pathlosses at heights above urban areas, where LOS conditions prevail.

## References


1. Saad W, Bennis M, Chen M (2020) A vision of 6G wireless systems: applications, trends, technologies, and open research problems. IEEE Netw 34(3):134–142. https://doi.org/10.1109/MNET.001.1900287
2. Ilchenko M, Kravchuk S (2020) Mobile infocommunication systems. Info Telecommun Sci 11(1):11–19. https://doi.org/10.20535/2411-2976.12020.11-19
3. Ilchenko MY, Kravchuk SO (2017) Telecommunication Systems. Naukova dumka, Kyiv. ISBN: 978–966–00–1566–1 (in Ukraine)
4. Namuduri K, Chaumette S, Kim JH, Sterbenz JPG (eds) (2017) UAV Networks and Communications. Cambridge University Press. ISBN: 9781107115309
5. Zgurovsky M, Ilchenko M, Kravchuk S, Kotovskyi V, Narytnik T, Cybulskyi L (2016) Prospects of using of aerial stratospheric telecommunication systems. In: Proceedings of the 2016 IEEE international scientific conference "RadioElectronics & InfoCommunications" (UkrMiCo 2016), Kyiv, Ukraine, 11–16 September 2016. IEEE Conference Publications, IEEE Xplore Digital Library. https://doi.org/10.1109/UkrMiCo.2016.7739636
6. Zakharov A, Rozenko S, Ilchenko M (2019) Varactor-tuned microstrip bandpass filter with loop hairpin and combline resonators. IEEE Trans Circ Syst II Exp Briefs 66(6):953–957
7. Valavanis KP, Vachtsevanos GJ (eds) (2015) Handbook of unmanned aerial vehicles. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-9707-1
8. Aerial base station. From wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Aerial_base_station
9. Li B, Fei Z, Zhang Y (2019) UAV Communications for 5G and beyond: recent advances and future trends. IEEE Internet Things J 6(2):2241–2263. https://doi.org/10.1109/JIOT.2018.2887086
10. Mozaffari M, Kasgari ATZ, Saad W, Bennis M, Debbah M (2019) Beyond 5G with UAVs: foundations of a 3D wireless cellular network. IEEE Trans Wireless Commun 18(1):357–372. https://doi.org/10.1109/TWC.2018.2879940
11. Vasiliev DS, Abilov A, Kaysina IA, Meitis DS (2018) Simulation-based assessment of quality of service in UAV-assisted mmWave system in crowded area. 23rd conference of open innovations association (FRUCT), Bologna, Italy, November, pp 13–16. https://doi.org/10.23919/FRUCT.2018.8588102
12. Al-Hourani A, Kandeepan S, Jamalipour A (2014) Modeling air-to-ground path loss for low altitude platforms in urban environments. 2014 IEEE global communications conference, 8–12 December 2014, pp 2898–2904. https://doi.org/10.1109/GLOCOM.2014.7037248
13. Ranjan A, Panigrahi B, Rath HK, Misra P, Simha A, Sahu HB (2018) A study on pathloss model for UAV based urban disaster and emergency communication systems. Twenty fourth national conference on communications (NCC), 25–28 February (2018), Hyderabad, India. https://doi.org/10.1109/NCC.2018.8600260
14. Mozaffari M, Saad W, Bennis M, Debbah M (2015) Drone small cells in the clouds: design, deployment and performance analysis. IEEE global communications conference (GLOBECOM), San Diego, CA, USA, 6–10 December (2015). https://doi.org/10.1109/GLOCOM.2015.7417609
15. Hayajneh AM, Zaidi SAR, McLernon D, Ghogho M (2016) Optimal dimensioning and performance analysis of drone-based wireless communications. IEEE global communications workshops (GC Wkshps), 4–8 December 2016, Washington, DC, USA. https://doi.org/10.1109/glocomw.2016.7848992
16. Mozaffari M, Saad W, Bennis M, Debbah M (2016) Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage. IEEE Commun Lett 20(8):1647–1650. https://doi.org/10.1109/lcomm.2016.2578312
17. Shi W, Li J, Xu W, Zhou H, Zhang N, Zhang S, Shen X (2018) Multiple drone-cell deployment analyses and optimization in drone assisted radio access networks. IEEE Access. 6:12518–12529. https://doi.org/10.1109/ACCESS.2018.2803788

18. Zulkifley MA, Behjati M, Nordin R, Zakaria MS (2021) Mobile network performance and technical feasibility of LTE-powered unmanned aerial vehicle. Sensors 21:2848. https://doi.org/10.3390/s21082848
19. Li X (2018) Deployment of drone base stations for cellular communication without Apriori user distribution information. IEEE 2018 37th Chinese control conference (CCC), 25–27 July 2018, Wuhan, China. https://doi.org/10.23919/ChiCC.2018.8482797
20. Kravchuk S, Kaidenko M, Afanasieva L, Kravchuk I (2020) Testing of the drone swarms as a communication relay system. Info Telecommun Sci 11(1):92–101. https://doi.org/10.20535/2411-2976.12020.92-101
21. Kravchuk SO (2019) Unmanned aerial vehicle as a mobile M2M-gateway for smart ubiquitous network. Digital of the 13th international scientific conference "Modern Challenges in Telecommunications", Kyiv, Ukraine. http://scholar.google.com/scholar?cluster=211967857 3849458940&hl=en&oi=scholarr (in Ukraine)
22. Kravchuk SO, Kravchuk IM (2020) Using RFID technology at operating a drone swarms in communication system mode. Info Telecommun Sci 12(2):16–23. https://ela.kpi.ua/bitstream/123456789/40551/1/ITS2020_11-2_p16-23.pdf
23. Kaidenko M, Kravchuk S (2019) Creation of communication system for unmanned aerial vehicles using SDR and SOC technologies. IEEE 2019 International conference on information and telecommunication technologies and radio electronics (UkrMiCo), 9–13 September, pp 1–4. (IEEE Xplore Digital Library) https://doi.org/10.1109/UkrMiCo47782.2019.9165422
24. Kaidenko M, Kravchuk S (2021) Autonomous unmanned aerial vehicles communications on the base of software-defined radio. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technology and Systems, vol 152. Lecture Notes in Networks and Systems. Springer, Cham, pp 289–302. https://doi.org/10.1007/978-3-030-58359-0_16
25. Ilchenko M, Kravchuk S, Kaydenko M (2019) Combined Over-the-Horizon Communication Systems. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technologies, vol 560. Lecture Notes in Electrical Engineering. Springer, Cham, pp 121–145. https://doi.org/10.1007/978-3-030-16770-7_6
26. Kravchuk S, Afanasieva L (2019) Formation of a wireless communication system based on a swarm of unmanned aerial vehicles. Info Telecommun Sci 1:11–18. https://doi.org/10.20535/2411-2976.12019.1-18
27. Ilchenko M, Kravchuk S, Minochkin D, Afanasieva L (2018) Troposcatter communication link model based on ray-tracing. Info Telecommun Sci 2:15–20. https://doi.org/10.20535/2411-2976.22018.15-20
28. Kravchuk S, Afanasieva L (2019) Wireless cooperative relaying without maintaining a direct connection between the source and target receiver terminals. Info Telecommun Sci 10(2):5–11. https://doi.org/10.20535/2411-2976.22019.5-11
29. Afanasieva L, Kravchuk S (2021) Wireless Systems with New Cooperative Relaying Algorithm. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technology and Systems, vol 152. Lecture Notes in Networks and Systems. Springer, Cham, pp 274–288. https://doi.org/10.1007/978-3-030-58359-0_15
30. Kravchuk S, Minochkin D, Omiotek Z, Bainazarov U, Weryńska-Bieniasz R, Iskakova A (2017) Cloud-based mobility management in heterogeneous wireless networks. Proc PIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments, Wilga, Poland, 7 August. 104451W. https://doi.org/10.1117/12.2280888
31. Ilchenko MY, Kravchuk SO (eds) (2019) Advanced in the telecommunications 2019: monograph, p 336. Kyiv. ISBN 978–617–7734–12–2
32. Afanasieva L, Minochkin D, Kravchuk S (2017) Providing telecommunication services to antarctic stations. Proceedings of the 2017 international conference on information and telecommunication technologies and radio electronics (UkrMiCo) 11–15 September 2017, Odessa, Ukraine. IEEE Conference Publications (IEEE Xplore Digital Library), https://doi.org/10.1109/UkrMiCo.2017.8095419

33. Zheng Y, Wang Y, Meng F (2013) Modeling and simulation of pathloss and fading for air-ground link of HAPs within a network simulator. Proc of IEEE International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Beijing, China, October
34. Meng YS, Lee YH (2011) Measurements and characterizations of air-to-ground channel over sea surface at C-band with low airborne altitudes. IEEE Trans Veh Technol 60(4):1943–1948
35. Kravchuk SO, Kravchuk IM (2021) Analysis of propagation loss models for ground-to-air radio links. Proceedings of the fifteenth international scientific and technical conference, 12–16 April 2021, Kyiv, Ukraine, pp. 173–175. http://conferenc.its.kpi.ua/proc/article/view/230685/229646 (in Ukraine)
36. Deng N, Haenggi M (2018) A novel approximate antenna pattern for directional antenna arrays. IEEE Wirel Commun Lett 7(5):832–835. https://doi.org/10.1109/LWC.2018.2829741
37. Yu X, Zhang J, Haenggi M, Letaief KB (2017) Coverage analysis for millimeter wave networks: the impact of directional antenna arrays. IEEE J Sel Areas Commun 35(7):1498–1512. https://doi.org/10.1109/JSAC.2017.2699098
38. Balanis CA (1996) Antenna theory: analysis and design. John Wily & Sons, New York
39. Kravchuk SO, Rystsova KI (2021) Base station antenna pattern approximation. Proceedings of the fifteenth international scientific and technical conference, 12–16 April 2021, Kyiv, Ukraine, pp. 176–179. http://conferenc.its.kpi.ua/proc/article/view/230687/229648 (in Ukraine)

# Principles of Constructing Communication and Control Systems Protected from the Effects of Jamming Attacks for Small-Sized Unmanned Aerial Vehicles




**Mykola Kaidenko and Serhii Kravchuk**

**Abstract** The research aimed at developing methods for creating high-tech countermeasures is presented. Vulnerabilities of small-sized unmanned aerial vehicles (UAVs) in terms of communication and control systems are considered. The classification of attacks and hindrances according to various criteria is given. A taxonomy of various types of known attacks that UAVs may be subject to is presented. Intentional interference can lead to a temporary loss of control of the unmanned vehicle, its incapacitation, interception of control of the vehicle, and, as a consequence, its misuse. It is noted that the security of UAVs is a major problem, especially in terms of cyber-attacks and radio channel jamming attacks, while there is currently no typical way to effectively counter them. A promising architecture of a communication and control system for a small-sized UAV is presented, its key features and implementation options are considered. The presented architecture is aimed primarily at protecting against jamming and radio channel simulation.

**Keywords** Unmanned aerial vehicles, Intentional interference · Communication system · Control of the vehicle · Wireless technologies

## 1 Introduction

Today, unmanned aerial vehicles (UAVs) are used in various fields such as monitoring, search, and rescue missions, commercial services, scientific research, agriculture, and military purposes. UAVs are useful in various fields, but safety has always been a major concern [1–3]. In addition, the UAV carries confidential information related to the privacy of people and causes serious problems if the data is used by

M. Kaidenko (✉) · S. Kravchuk (✉)
Telecommunication Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Industrialnyi Lane 2, Kyiv 03056, Ukraine
e-mail: kkk610@ukr.net

S. Kravchuk
e-mail: sakravchuk@ukr.net

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_20

intruders. There have been many attempts to address the security issues of UAVs, but there is no typical way to protect UAVs from various attacks.

An important aspect of attacks on UAVs, including their abductions, which should be emphasized, is a targeted attack, including a cyberattack in terms of cyberwar terminology [4], which encompasses exploitation, attack, defense, and asymmetric response in general. An attempted hijacking is a combination of cyber and electronic operations/efforts, not a one-click hijacking, as it requires several tools and different methods from the fields of cyber warfare and electronic warfare simultaneously in a series of actions to achieve its goal, especially when the target is unauthorized using a drone for terrorist purposes, or when it is a military UAV. The multistage nature of modern attacks on UAVs, based on the coordination of individual actions [5], including a fully executed attack with UAV capture, or information interception. At the same time, it is important to know all the vulnerabilities of the unmanned vehicle and the possible actions of intruders when attacking it. As a form of a cyber-physical system, UAVs and the computing infrastructure that supports it are vulnerable to cyber-attacks on their hardware, software, communications, and data.

In addition to radar visibility, an important unmasking feature for carrying out an attack on a UAV is the presence of its own electromagnetic radiation from communication equipment, through which a data transmission channel is organized to the UAV flight control point or other consumers. Since the obtained information is communicated to consumers in real-time, this feature is the most significant and is fundamentally irreducible, which makes it possible to detect UAVs even by passive radio monitoring means.

## 2 Types of Attacks on UAVs and Threats Posed by Them

The main victims of unmanned aerial vehicle security breaches are various system components such as devices, networks, and communication channels [6]. In this case, the classification of attacks can be carried out according to various criteria.

All attacks on UAVs at the place of application can be conditionally divided into:

- attacks that are carried out directly on the UAV through wireless communication, control, and telemetry channel;
- attacks carried out on UAVs through the control network;
- attacks carried out on the UAV control network;
- attacks carried out on UAV software.

According to the time of implementation, attacks on UAVs are divided into:

- attacks in real-time;
- delayed attacks.

Attacks on UAVs can be divided into:

- attacks are aimed at the physical destruction of the UAV;
- attacks are aimed at intercepting control of the UAV for the purpose of unauthorized use or theft;
- attacks are aimed at intercepting information about the UAV and target information for which the unmanned vehicle is used, for example, video, data from sensors.

The types of attacks used by an attacker may differ significantly depending on the intended use of the UAV and the way it is controlled:

- an autonomous UAV with manual control by an operator from an autonomous control panel;
- an autonomous UAV with automatic flight control on the instructions of an autonomous control panel;
- autonomous UAV with network control using wireless technologies and networks;
- UAV Network (a swarm of drones) with automatic flight control, including network control.

Threats to an unmanned vehicle and useful information for the collection and processing of which it is used are responsible for breaches of security and lead to large losses in terms of resources, trust, and availability [7]. Based on these vulnerabilities, attacks and their impact can be categorized into five main types. In Fig. 1 shows a taxonomy of various types of known attacks on unmanned aerial vehicles, which can be applied depending on the target use of the vehicle, its type and class, and the objectives of the attack [8], generalized for all cases of using UAVs.

According to studies [9], the greatest threats of attacks on the UAV itself are: Fly Away, Resource Leak, Loss of GPS, Battery Depletes, Loss of Data Link, Fuel Depletes, Crash, Loss of Situational Awareness, Autopilot Software Error / Fail, Loss of Direct Visual, ground control station Failure, Hazard Weather, Automatic Transmission Locked, Hostile Environment.

## 3 Key Features that Determine the Architecture of Small-Sized UAV Communication Channels Protected from Intentional Interference

### *3.1 Intentional Interference on UAV Communication Channels and Ways to Combat Them*

One of the most vulnerable types of attacks on UAVs from the point of view of security and survivability is channel jamming attacks, and in a broader sense - attacks of destructive influence on the radio channel. Therefore, when building a communication and control system for a UAV, it is necessary to take into account the features of such attacks and minimize their impact on the survivability of an unmanned vehicle.

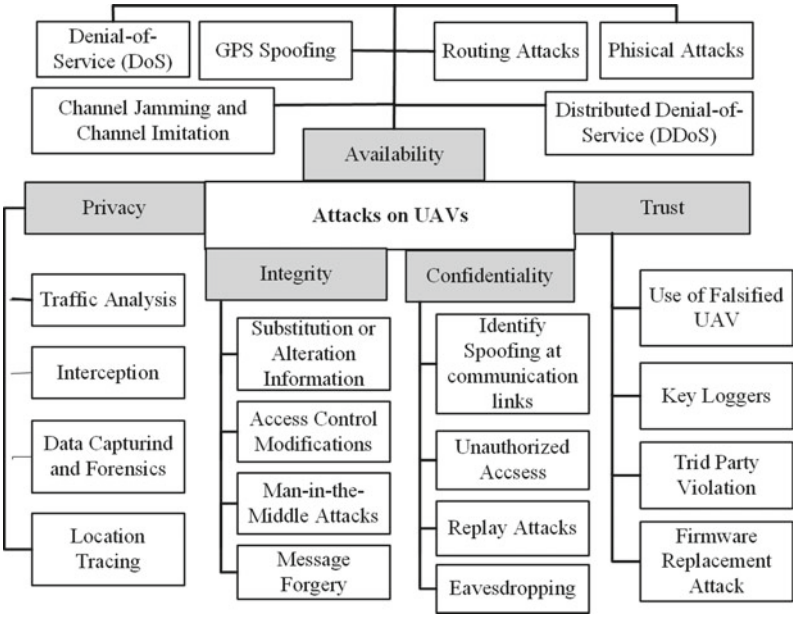



**Fig. 1** Taxonomy types of known attacks on the UAV (adapted from [8])

This is due to the fact that at the present stage of technology development, the market for inexpensive means of destructive influence on the radio channel is rapidly developing. Currently, the following methods of destructive influence on the radio channel, in particular on the UAV radio channel, are being determined:

- radio-electronic jamming by deliberate interference of transmitted data in the radio channel;
- interception of data transmitted in the radio channel;
- viewing data transmitted in the radio channel;
- substitution of the transmitted data in the radio channel.

Electronic jamming leads to a complete loss of control of the unmanned vehicle. Interception, viewing, and replacement of data transmitted in the radio channel can lead to the interception of control of the unmanned vehicle with the aim of kidnapping, misuse, or destruction, while the survivability of the UAV is affected by the substitution of data in the control channel. The task of protecting data from viewing is solved by using cryptographic methods of data protection, while for civilian UAVs, the length of the encryption key is limited. To intercept UAVs, viewing and completely substituting data is not a prerequisite; it is sufficient to use more power of previously intercepted and recorded data for transmission as a deliberate interference, which leads to a loss of control of the UAV. The most vulnerable from the point of view of interception and data spoofing are WiFi-based control channels for unmanned aerial

vehicles [10–13], although channels with proprietary control algorithms for UAVs cannot be completely secure, especially for civilian UAVs.

Certain works, including by Russian researchers [14], are devoted not to methods of protecting UAVs from intentional interference, but, on the contrary, to methods of countering unmanned aerial vehicles by creating, among other things, various types of intentional interference. At the same time, the main emphasis is placed not just on a cyberattack on a UAV, but on suppressing a communication channel or intercepting an unmanned vehicle by replacing information in the control channel.

In [15, 16], the architecture of the communication system of an unmanned vehicle was proposed to increase its survivability, which is achieved through the use of adaptation over the frequency range using an additional reception channel for continuous analysis of the interference situation, using two or more data transmission channels, complex algorithms of optimal selection of the operating range, transmission channel parameters and adaptive protocols for simultaneous data transmission over two or more channels. However, such a solution cannot always be used to protect small-sized UAVs, since the main limitation for the use of such a system is the limitation on the payload weight. This limitation leads to the fact that the number of onboard equipment should be minimal. At the same time, the use of two transmission channels requires twice as much energy, which leads to the need to increase the capacity of the storage battery.

Our proposed solution to this problem is the use of two data transmission channels in the UAV control channel and one channel in the telemetry channel. That is, in the direction from the control station to the UAV (Uplink), two channels are used, and in the direction from the UAV to the control station, one channel (Downlink). This approach is explained by the fact that suppression in the control channel is more easily realized due to the fact that the suppression station is most often located in the immediate vicinity of the UAV mission, which in this case is guaranteed to be in the line of sight. Significant interference can be created by a simulated signal with lower power at the input of the UAV receiver than the useful signal. The use of directional antennas and low-power transmitters ensures the secrecy of the use of the electronic warfare station. In the telemetry channel, the situation is absolutely opposite, when using antennas aimed at the UAV, the power emitted by the EW station should significantly exceed the radiation power of the useful signal by the UAV transmitter, in addition, the UAV control station and the EW station are most often not in the line of sight.

## *3.2 Ranges of Operating Frequencies and Modes of Operation of the UAV Communication System*

The use of two or more channels for control and telemetry in order to counteract the influence of deliberate interference complicates the architecture of the UAV communication system, while the possibilities for using various combinations of

frequency bands are limited by permits. In this case, both TDD and FDD modes can be used.

For civilian UAVs, the frequency ranges for control and telemetry channels are fixed from the list of permitted civilian use. Such frequency ranges can be [17]:

- 433 MHz telemetry and radio remote control range (433.05–434.79 MHz) with 1.74 MHz frequency band with support for TDD mode only;
- 868 MHz telemetry and radio remote control range (868–868.6 MHz) with 0.6 MHz frequency band with support for TDD mode only;
- range of wireless phones 814/904 MHz (814–815; 904–905 MHz) with a paired frequency band up to 1 MHz with support for FDD mode;
- 2.4 GHz band (2.4–2.483 GHz) with a frequency band of 83 MHz with traditional TDD mode support, but FDD mode is also possible;
- 5.8 GHz range, (5.150–5.350; 5.470–5.850 Hz) with frequency bands up to 100 MHz with traditional TDD support, but FDD is also possible.

It should be noted that in the European Table of Frequency Allocations and Annexes [17], the permitted frequencies for civil use may differ for national use. For military use, this document also specifies frequencies for the military use of UAVs, for example, the paired frequency range of 874.4–880.0 MHz and 919.4–925.0 MHz, and the possibilities for using different frequency ranges are much greater.

At first glance, the use of the TDD mode is the most acceptable from the point of view of implementation and minimization of the occupied bandwidth. However, from the point of view of increasing the survivability of the UAV communication channel under conditions of deliberate interference, it is preferable to use the FDD mode. This is due to the fact that in order to set up concentrated simulation interference, the radio reconnaissance station, which is most often in the zone in the area of the UAV mission, must determine the parameters of the UAV communication channel, which in the TDD mode makes it possible to simultaneously determine the parameters of the control channel and the parameters of the telemetry channel. In the case of the FDD mode, an independent definition of the parameters of the control channel and the telemetry channel is required.

When using two frequency ranges for the organization of control channels, options are possible with operation on one antenna or two antennas. The most optimal option is to use one wide-band (or dual-band) antenna in a circular directional pattern for the UAV and two spaced band antennas at the control station. The advantage of using one antenna is its weight and size while operating on one antenna increases the requirements for the filtering system to suppress out-of-band interference between channels. The advantage of dual antenna operation is that there is no correlation of channel fading due to both frequency diversity and space diversity, which provides more robust communication in the presence of frequency selective fading.

For efficient operation of the UAV control system when using two channels in two frequency ranges, the power in both channels must be the same, in this case, the wireless control channel can be perceived as a dual-frequency MIMO-channel. The equalization of the power in the channel is necessary both to maintain the same energy potential in both channels and to ensure the admissible energy secrecy of

the communication channel. It can also be used as one of the criteria for detecting simulated interference. As is known from [18], the attenuation of a radio wave in free space is:

$$L_{bf} = 20log\left(\frac{4\pi d}{\lambda}\right), \tag{1}$$

where $L_{bf}$—free-space basic transmission loss (dB); $d$—distance; λ—wavelength, and $d$ and λ are expressed in the same unit.

Equation (1) can also be written using the frequency instead of the wavelength

$$L_{bf} = 32,4 + 20\log(f) + 20\log(d), \tag{2}$$

where $f$—frequency (MHz), $d$—distance (km).

As can be seen from expression (2), the damping depends on the frequency in direct proportion. Thus, in the first approximation, it is sufficient to radiate constant power in both channels to ensure the same energy potential of the radio link. In a real system, it is necessary to take into account the gain of the antennas (antennas when transmitters operate on one antenna) and the sensitivity of the receivers at different frequencies. Taking into account the fact that in the case of using multi-band antennas, the antenna gains increase with increasing frequency with unchanged dimensions, and the noise figure of the input path varies within 1 dB, the output power in both channels will be of the same order. In addition, in order to maintain the same energy potential of radio links, the telemetry channel additionally transmits the measured level of the input signal RSSI in both channels and the measured signal-to-noise ratio in both channels SNR (Signal-to-noise ratio).

## *3.3 Evaluation of the Survivability of the UAV Communication Channel and the Advantages of Working in Two Frequency Ranges*

The survivability of a communication channel is defined as the ability to work under conditions of exposure to various kinds of destabilizing factors in the form of unintentional and deliberate interference. Methods for dealing with unintentional interference are widely known and survivability can be indirectly determined through the probability of a bit error, or the probability of an error in a data block. In the case of intentional interference, this approach to determining survivability cannot be applied due to the very nature of the intentional interference. A more accurate approach for determining survivability may be the probability of maintaining the operating parameters of the control channel when exposed to deliberate interference.

For the proposed architecture of the communication and control system of a small-sized UAV, a formula of complex probability can be applied depending on

the occurrence of events. It takes into account the following variables, assuming that a deliberate interference in the form of imitation interference acts on the UAV control channel: the probability of the radio countermeasure system operation in the frequency ranges used; the likelihood that there are two countermeasures that can work at the same time; the probability that two countermeasure systems can work not only simultaneously, but also be synchronized in time, the probability of correct detection and false alarms in the presence of intentional imitation interference.

The survivability factor of the UAV control channel in the case of using one channel and affecting it only from unintended interference can be defined as $K_{liv} = 1—P_{ber}$, where $K_{liv}$—control channel survivability factor; $P_{ber}$—communication channel bit error probability.

The survivability factor of the UAV control channel in the case of using two channels and the impact on them only unintentional interference can be defined as.

$K_{liv} = 1—P_{ber_chan1}\ P_{ber_chan2}$, where $P_{ber_chan1}$ and $P_{ber_chan2}$—bit error probabilities in channels, events are independent.

The survivability coefficient of the UAV control channel in the case of using one channel and the impact on it of both unintentional interference and intentional jamming interference can be defined as: $K_{liv} = 1—P_{ber}/P_{jamm}$; where $P_{jamm}$—the probability of suppression of the communication channel by jamming interference.

The survivability coefficient of the UAV control channel in the case of using two channels and the impact on them of both unintentional interference and intentional jamming interference, or imitation from two independent electronic warfare systems can be defined as:

$$K_{liv} = 1 - \frac{P_{ber_chan1}}{P_{jamm_chan1}} \cdot \frac{P_{ber_chan2}}{P_{jamm_chan2}},$$

where $P_{jamm_chan1}$, $P_{jamm_chan2}$—the probabilities of suppression of the communication channel by the jamming (imitation) noise of the event are independent.

The survivability coefficient of the UAV control channel in the case of using two channels and the impact on them of both unintentional interference and intentional imitation interference can be determined as:

$$K_{liv} = 1 - \frac{P_{ber_chan1}}{P_{im_chan1}} \cdot \frac{P_{ber_chan2}}{P_{im_chan2}} \cdot P_{sync}, \tag{3}$$

where $P_{im_chan1}$, $P_{im_chan2}$—probabilities of suppression of a communication channel by imitation interference, events are independent; $P_{sync}$—the probability of synchronous operation of the imitation jamming system in two channels.

As can be seen from expression (3), the survivability of the UAV control channel will always be high. Only in the case of using a complex radio countermeasure system operating simultaneously in two frequency ranges with the possibility of setting up a synchronous imitation jamming, the survivability of the control channel will be significantly reduced.

The use of two frequency ranges, in addition to increasing the survivability of the UAV control channel in case of exposure to deliberate interference, also provides a greater channel resistance to multipath fading [19]. To qualitatively evaluate the efficiency of using two frequency bands to mitigate the effects of multipath fading, the frequency response of multipath channels for Rice and Rayleigh fading was simulated. The results of multipath fading simulations are shown in Figs. 2, 3, 4, 5. The modeling was carried out in the baseband at a UAV speed of 108 km/h (30 m/s), using linear QPSK modulation, pre-modulation filtering using a square-root-raised-cosine filter with a rounding index of 0.2 and decimation of 4. The channel with Rice fading was modeled by three beams with delays of 167 and 350 ns (attenuation of 3- and 6-dB beams), such fading characteristics are typical for reflection from the earth's surface at a distance to the UAV of 10 km, located at an altitude of 1 km. The channel with Rayleigh fading was modeled by five beams with delays and attenuation of the beams typical for a distance to the UAV of 10 km located in the urban area. Fading was simulated in two frequency ranges for the case of a low-speed channel (38.4 kbps) with the FDMA multiple access modes (frequency bands 433 and 860 MHz) and for the case of a high-speed channel (33.4 Mbps) with the TDMA multiple access modes (frequency ranges 2400 and 5800 MHz). The simulation results of the low-speed channel are shown in Figs. 2, 3. As can be seen from the figures, in the case of the Rice channel, fading does not differ much for different frequency ranges (different Doppler shifts); greater noise immunity in one of the channels compared to the other. The simulation results of a high-speed channel are shown in Figs. 4, 5. As can be seen from the figures, in the case of Rice fading, there is a shift in the points of the minimum of fading, and in a channel with Rayleigh fading, the fading characteristic is significantly different, with a significant spread of the spectrum for the upper-frequency range.

### *3.4 Features of the Protocol for Determining Jamming Interference and Simulation Interference*

Simulated structured interference detection algorithms are used to combat channel simulation interference. These algorithms are based on channel synchronization, maintaining a given level of channel energy and monitoring energy characteristics in channels, marking data packets. Maintaining a given level of channel energy, which should be the same in both channels, provides additional secrecy for the transmission of information, reducing the likelihood of interception by the station, which is not in the line of sight. Channel synchronization and maintenance of a given energy level are realized by using software-defined radio and systems-on-a-chip technologies described in [20]. The available measured channel parameters for detecting interference are the signal-to-noise ratio, the received signal level, and the change in their behavior over time, measured as probability distribution density.

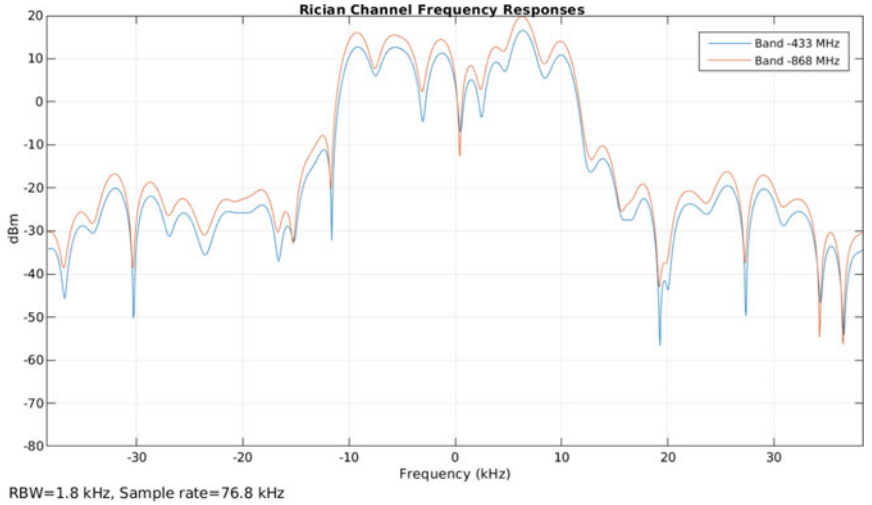



**Fig. 2** Frequency response of multipath fading with Rice distribution in a low-speed UAV control channel for the frequency ranges 433 and 860 MHz

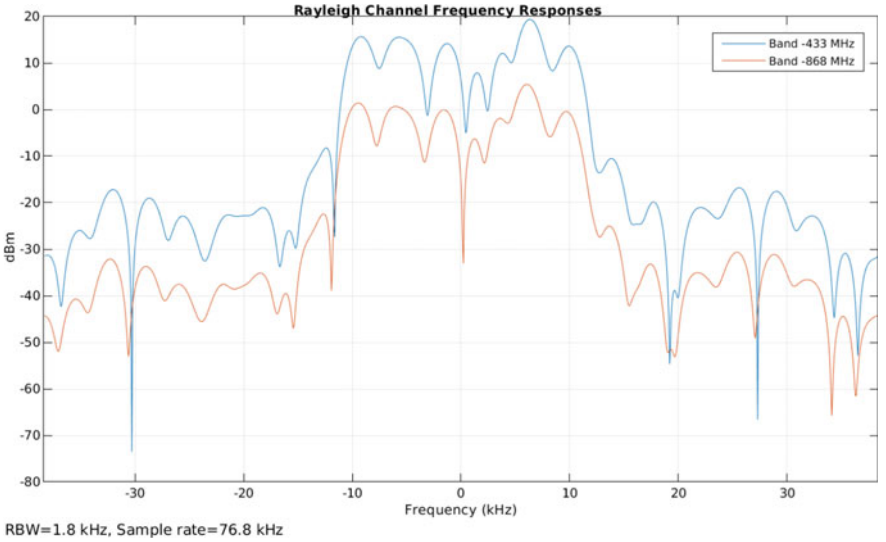



**Fig. 3** Frequency response of multipath fading with Rayleigh distribution in a low-speed UAV control channel for the frequency ranges 433 and 860 MHz

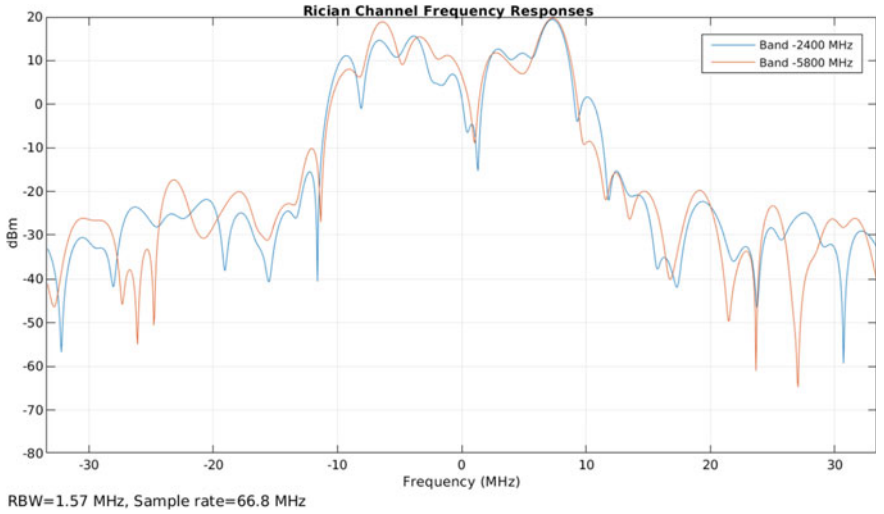



**Fig. 4** Frequency response of multipath fading with Rice distribution in a high-speed UAV control channel for the frequency ranges of 2400 and 5800 MHz

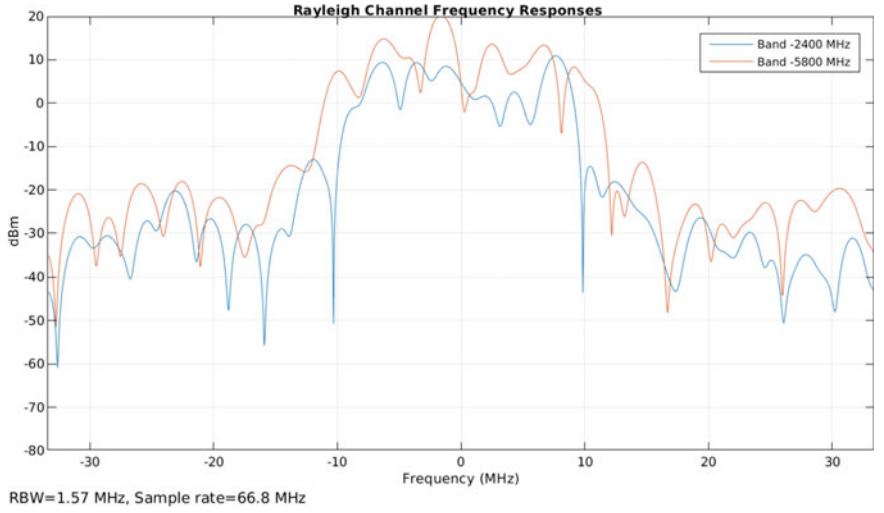



**Fig. 5** Frequency response of multipath fading with Rayleigh distribution in a high-speed UAV control channel for the frequency ranges of 2400 and 5800 MHz

Evaluating the impact of jamming interference on a control channel is straightforward, since a significant reduction in signal-to-noise ratio and a measured degradation of the bit error rate can be a reliable estimate of jamming in a channel, and these events for the two channels may be uncorrelated.

In the case of using channel simulation for interception or interference, the algorithm for determining the interference is much more complicated due to the fact that it is necessary to determine, firstly, whether the interference was imitated and, secondly, in which channel it was carried out.

In the protocol for determining imitation noise, fuzzy inference algorithms are used, similar to those proposed in [21] and used in [22, 23], in which the rules of fuzzy productions in the general form are applied, presented as.

$$\text{RULE_i} < \# > : \text{IF}''\text{EVENT_i}''\text{THEN}''\text{INFERENCE_i}'', \forall \text{i} = 1 : \text{N}, \quad (4)$$

where N—the total number of terms.

Defasification (reduction to clarity) is carried out for each output separately. The decision on the presence of imitation noise in the channel is made by majority vote according to all the rules using weight coefficients for each conclusion (4).

The use of the fuzzy inference algorithm is associated with the difficulty of reliably determining the imitation noise. The task of the algorithm in the operation of two reception channels, in which the same information is transmitted, is to determine in which of the channels the information comes from the ground station, and in which from the source of intentional interference. It should be noted that the simulation interference is effective for complete control interception of the UAV only in the case of complete imitation of the information message with full repetitions of the information message, against imitation inserts and a cryptographic key. Without these conditions, the simulation noise becomes similar to the jamming noise, except that the channel jamming noise suppresses the channel with a powerful noise signal, usually not directed to a specific communication channel, and the simulation noise can jam the channel with a signal comparable in power to the useful signal, violating the synchronization system. both at the physical layer and at the MAC layer. In this case, when the power of the interference signal exceeds the main signal, the UAV receiver at the physical level perceives the mixture of the signal with interference as a multipath channel. Due to this, the effectiveness of the system for suppressing the communication channel with imitation noise (and, accordingly, the complexity of countering it) is much higher than the efficiency of the jamming system with a powerful signal due to the fact that it can be relatively low-power, difficult to determine both from the point of view of the information component and from the point of view of determining the location of the source of interference.

## 4 the Structure of the Communication and Control System of Small-Sized UAVs

Taking into account the features that affect the architecture of building communication systems for small-sized UAVs, variants of the structure of the onboard equipment of the UAV and the ground control station were developed.

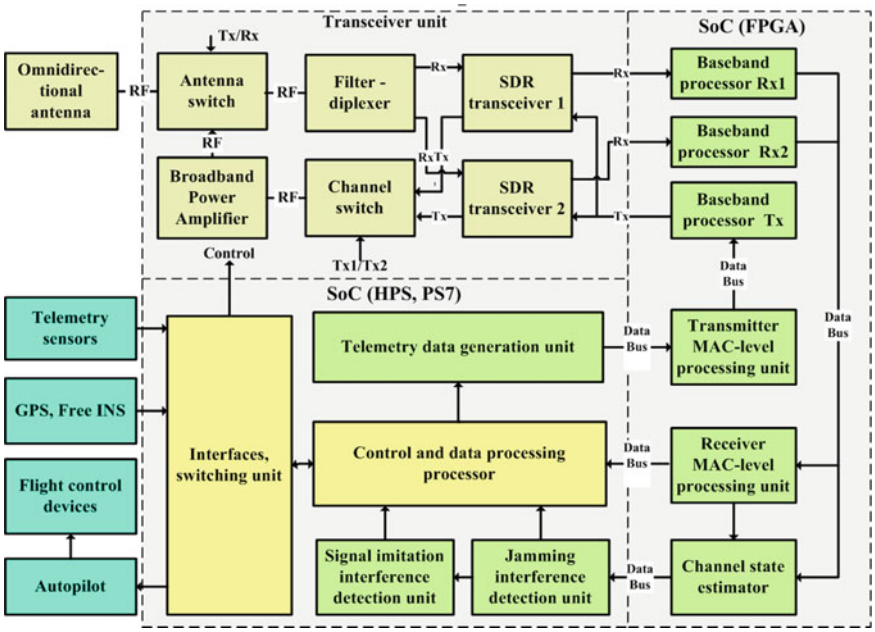



**Fig. 6** The structure of the onboard equipment of the communication and control system of a small-sized UAV in TDD mode

Figure 6 shows the structure of the onboard equipment of the communication and control system of a small-sized UAV, which operates in the TDD mode.

The main modules are highlighted in the structure: Transceiver unit; a module for processing the baseband and lower MAC level, implemented on SoC FPGA [20], and a module for interfaces and processing information flows, implemented on SoC HPS (Intel) or SoC PS7 (Xilinx) [20].

The transceiver unit implements the RF part and is built on the basis of two identical SDR transceivers (eg AD9364 [24]), which must be synchronized from the same reference signal source for the effective operation of the deliberate interference detection system. The UAV uses a dual-band antenna with a circular directional pattern. Antenna switch is used to separate transmission and reception modes. Each of the SDR transceivers is tuned to operate in its own frequency range, the division of the frequency ranges in the receive mode is carried out in the Filter-diplexer. In transmission mode, the choice between transmission frequencies (bands) is performed by Channel swithc. The circuit uses a Broadband Power Amplifier, which reduces power consumption and weight, and dimensions. The use of two frequency ranges for transmission is due to the fact that in the case of setting up a powerful deliberate jamming interference, the telemetry channel will also be affected by this interference.

In SoC FPGA unit, baseband processing functions are implemented, including mapping, demapping, FEC (coding and decoding), interliving, pulse-shaping ect. Two Baseband processors are used for receiving and one for transmitting, and the receiving Baseband processors can be implemented as a single physical module operating at a double clock frequency and equipped with a memory buffer to store input data from both receive channels. The baseband processor data is fed in parallel to both SDR transceivers to ensure minimum band hopping times. It also implements the lower MAC-level processing functions for the Transmitter & Receiver and the Channel state estimation function, which estimates the noise immunity of the received signal on both channels.

The SoC HPS module is implemented on the SoC processor running the Embedded Linux operating system, it performs all the control functions of the SDR transceivers and Transceiver unit. In addition, it implements the upper MAC-level processing functions for the Transmitter & Receiver in the Control channel data generation unit and Control and data processing processor unit. This module implements algorithms for detecting intentional interference in the Jamming interference detection unit and Signal imitation interference detection unit, information for the operation of which uses Channel state estimation, received RSSI signal levels measured in SDR transceivers, as well as estimates that are carried out at the MAC-level. Since the SoC HPS is implemented on an ARM processor, it performs the functions of switching and interface with external devices. In particular, these are GPS and Free INS (if it is installed to increase the survivability of the UAV), Telemetry sensors, Flight control devices, and Autopilot.

In Fig. 7 shows the structure of the equipment of the ground station of the communication and control system of a small-sized UAV, which operates in TDD mode. The ground station uses an antenna system with automatic tracking of the direction to the UAV Antenna system with UAV automatic tracking system, which uses directional antennas (or one two-band antenna to minimize dimensions) to optimize the energy potential of the radio link (this is very important for the channel from the UAV to the ground Downlink Channel station). The difference from the structure of the UAV onboard equipment is that since two frequency ranges are used simultaneously for transmission, it is necessary to use two Antenna switches for each channel. When working on one antenna, a Filter–diplexer is used, and when working on two antennas, two Bandpass filters. The structure shows the Power Amplifier Unit, which is two independent power amplifiers, which can be identical broadband. It is worth noting that power control in both the ground station and the UAV is carried out in the SDR transceiver. In the receive path, the receivers of both SDR transceivers are used with switching between channels already in the Baseband processor Rx, this is necessary for fast switching from band to band.

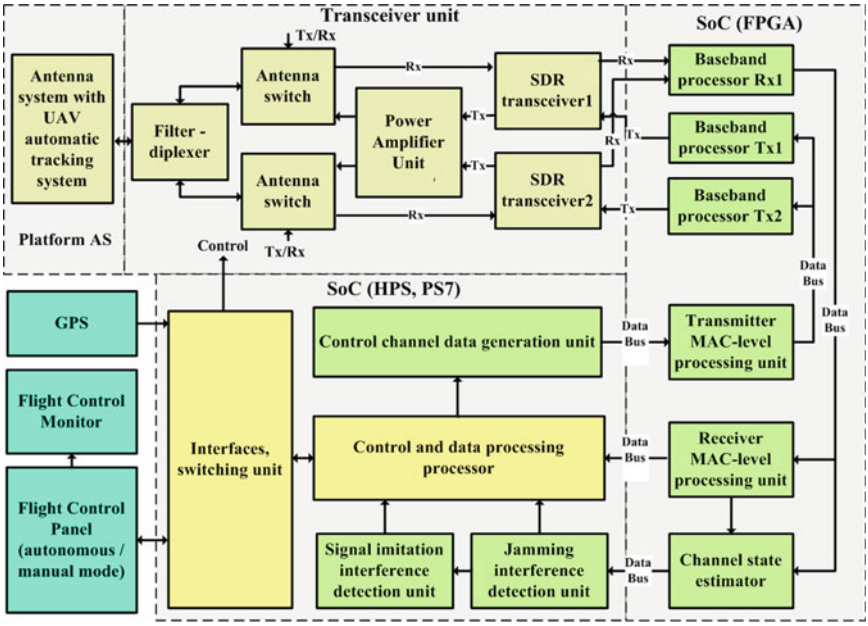



**Fig. 7** The structure of the equipment of the ground station of the communication and control system of a small-sized UAV in the TDD mode

The next difference is the Interfaces and switching unit of the SoC HPS module, which performs the functions of switching and interfaces with external devices of the ground station: Flight Control Monitor, Flight Control Panel (works both autonomous and in manual mode), and a GPS receiver. The presence of a GPS receiver in a ground station is due to the need to synchronize with the UAV at the initial stages of flight and to minimize the effect of interference such as GPS Spoofing.

Figures 8, 9 and 10 show variants of the structure of the onboard equipment of the communication and control system of a small-sized UAV in the FDD mode. Since the changes compared to the structure of the equipment for the TDD mode concern only the Transceiver unit, the structure of these particular modules is presented. In the structure shown in Fig. 8, only one frequency range is used in the telemetry channel; this is the simplest structure, the disadvantage of which is the ability of an attacker to suppress the telemetry channel. At the same time, its advantage is a simpler filtration system and the use of a Power Amplifier, designed to work in only one range. Such a structure can be used if it is not possible to use two paired frequency ranges to build a UAV communication system.

If it is possible to use two paired frequency ranges, it becomes possible to work either in one range or in another. In this case, two variants of the structure are possible: with the use of two fully duplicated transmission tracks with two Power Amplifiers

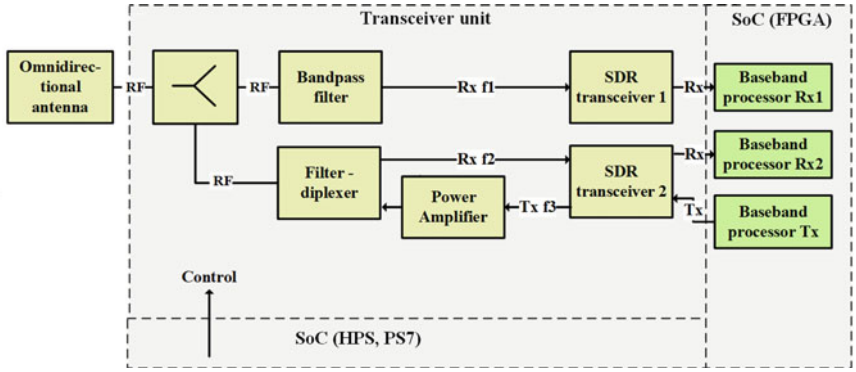



**Fig. 8** The structure of the onboard equipment of the communication and control system of a small-sized UAV in the FDD mode using one frequency in the telemetry channel

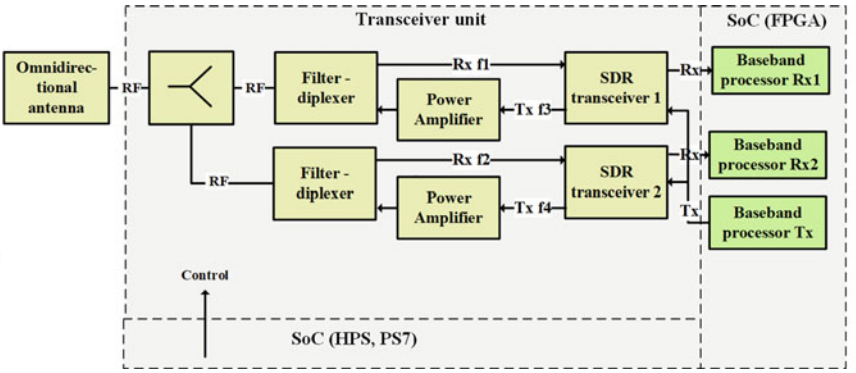



**Fig. 9** The structure of the onboard equipment of the communication and control system of a small-sized UAV in the FDD mode using two frequencies in the telemetry channel and two power amplifiers

and a Filter-diplexer (Fig. 9), or using a Broadband Power Amplifier, two Filter-diplexers and two Channel switches (Fig. 10). The second option is more compact in terms of dimensions.

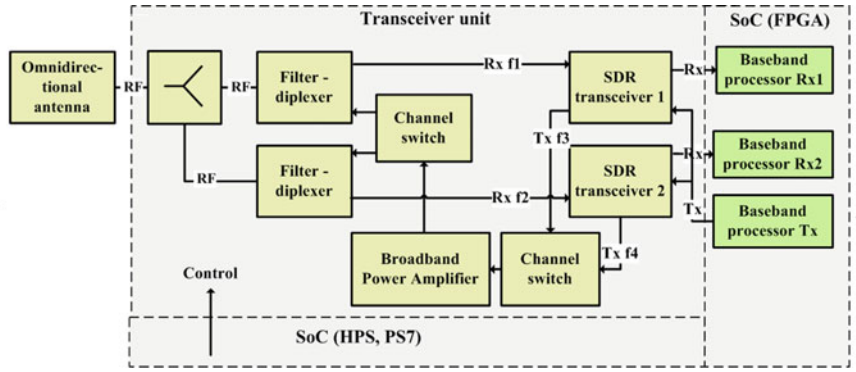



**Fig. 10** The structure of the onboard equipment of the communication and control system of a small-sized UAV in the FDD mode using two frequencies in the telemetry channel and one power amplifier

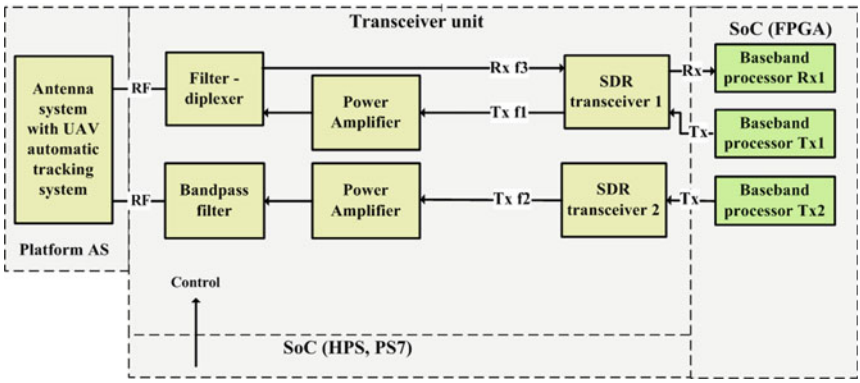



**Fig. 11** The structure of the equipment of the ground station of the communication and control system of a small-sized UAV in the FDD mode using one frequency in the telemetry channel

Figures 11, 12 show the structure of the ground station equipment for the communication and control system of a small-sized UAV in the FDD mode. In this structure, depending on the number of frequency ranges available for use in the telemetry channel, either one SDR transceiver receiver (Fig. 11) or two simultaneously (Fig. 12) is used channels in both bands have identical characteristics.

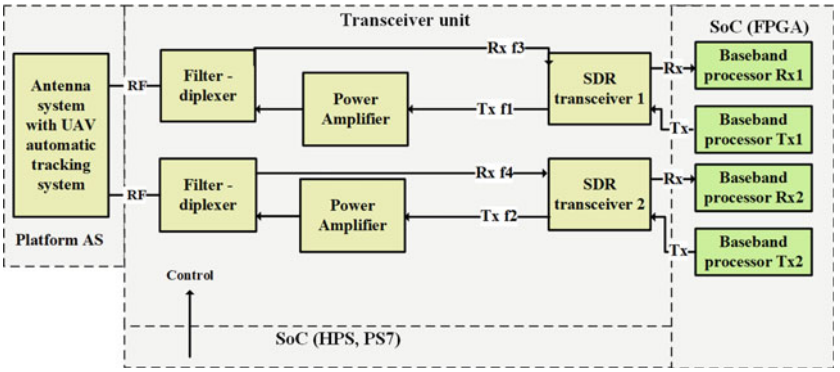



**Fig. 12** The structure of the equipment of the ground station of the communication and control system of a small-sized UAV in the FDD mode using two frequencies in the telemetry channel

## 5 Conclusion

The paper presents the results of work on the creation of communication and control systems for small-sized UAVs that provide protection against the effects of deliberate interference on them, namely, jamming and imitation.

A classification of types of attacks on UAVs and the threats posed by them is presented, as well as a taxonomy of known types of attacks on UAVs. It has been determined that the most vulnerable types of attacks on UAVs from the point of view of safety and survivability are attacks such as channel jamming, and in a broader sense—attacks of destructive influence on the radio channel.

An approach is proposed to combat intentional interference by using two data transmission channels in the UAV control channel and one channel in the telemetry system. The proposed protection solutions using two frequency ranges for the UAV control channel, on the one hand, significantly increase the survivability of the communication system of the vehicle, and on the other hand, provide a higher channel resistance to multipath fading. To qualitatively evaluate the efficiency of using two frequency bands to mitigate the effects of multipath fading, the frequency response of multipath channels for Rice and Rayleigh fading was simulated.

The use of two or more channels for control and telemetry in order to counteract the influence of deliberate interference complicates the architecture of the UAV communication system, while the possibilities for using various combinations of frequency bands are limited by permits. In this case, both TDD and FDD modes can be used. The use of the TDD mode is the most acceptable from the point of view of implementation and minimization of the occupied bandwidth. However, from the point of view of increasing the survivability of the UAV communication channel under conditions of deliberate interference, it is preferable to use the FDD mode.

The key features of the protocol (algorithm) of the detector of intentional jamming and imitation, which is based on the use of the mathematical apparatus of fuzzy inference, are presented. The use of fuzzy inference is associated with the difficulty of

reliably determining the imitation noise. The task of the algorithm in the operation of two reception channels, in which the same information is transmitted, is to determine in which of the channels the information comes from the ground station, and in which from the source of intentional interference.

Taking into account the results of the studies carried out, the structure of the communication system of a small-sized UAV and a ground control station was developed for various modes of operation, depending on duplex technologies, used antenna systems, available for using radio frequency bands. The presented structure is focused on implementation using SDR and SoC technologies.

Further work will be aimed at creating and testing a real communication system for a small-sized UAV, protected from the effects of deliberate interference of various nature, including jamming and imitation interference.

## References

1. Namuduri K, Chaumette S, Kim JH, Sterbenz JPG (eds) (2017) UAV Networks and Communications. University Press, Cambridge. ISBN: 9781107115309
2. Kravchuk S, Kaidenko M, Afanasieva L, Kravchuk I (2020) Testing of the drone swarms as a communication relay system. Info Telecommun Sci 11(1):92–101. https://doi.org/10.20535/2411-2976.12020.92-101
3. Kravchuk S, Afanasieva L (2019) Formation of a wireless communication system based on a swarm of unmanned aerial vehicles. Information and Telecommunication Sciences 1:11–18. https://doi.org/10.20535/2411-2976.12019.11-18
4. Giray S (2013) Anatomy of unmanned aerial vehicle hijacking with signal spoofing. 2013 6th international conference on recent advances in space technologies (RAST), Istanbul, Turkey, pp. 795–800. https://doi.org/10.1109/RAST.2013.6581320
5. Yaacoub J, Noura H, Salman O, Chehab A (2020) Security analysis of drones systems: attacks, limitations and recommendations. Internet Things 11:1–39. https://doi.org/10.1016/j.iot.2020.100218
6. Yampolskiy M, Horvath P, Koutsoukos XD, Xue Y, Sztipanovits J (2013) Taxonomy for description of cross-domain attacks on cps. Proceedings of the 2nd ACM international conference on high confidence networked systems, pp. 135–142. ACM, Philadelphia, Pennsylvania, USA
7. Mansfield K, Eveleigh T, Holzer TH, Sarkani S (2013) Unmanned aerial vehicle smart device ground control station cyber security threat model. Proceedings of the IEEE conference on technologies for Homeland Security (HST'12), Waltham, MA, USA, November. https://doi.org/10.1109/THS.2013.6699093
8. Choudhary G, Sharma V, Gupta T, Kim J, You I (2018) Internet of Drones (IoD): Threats, Vulnerability, and Security Perspectives, The 3rd International Symposium on Mobile Internet Security (MobiSec'18), Auguest 29-September 1, 2018. Cebu, Philippines, 37:1–13
9. Jafary B, Bhattacharya S, Nafreen M, Yuan S, Zhou J, Wu L, Manjunath P, Chigan T, Fiondella L (2019) The application of unmanned aerial systems in surface transportation, volume II-F: Drone cyber security: assurance methods and standards, report no 19–010, massachusetts department of transportation office of transportation planning, Boston, MA 02116, 76. https://www.researchgate.net/publication/340539206
10. Sedjelmaci H, Senouci SM (2018) Cyber security methods for aerial vehicle networks: taxonomy, challenges and solution, J. Supercomput, pp 1–17. https://doi.org/10.1007/s11227-018-2287-8

11. Sedjelmaci H, Senouci SM, Ansari N (2017) A hierarchical detection and response system to enhance security against lethal cyber-attacks in uav networks. IEEE Trans. Syst. Man Cybernet 48(9):1594–1606
12. He D, Chan S, Guizani M (2017) Drone-assisted public safety networks: the security aspect. IEEE Commun Mag 55(8):218–223
13. Westerlund O, Asif R (2019) Drone hacking with Raspberry-Pi 3 and WiFi pineapple: security and privacy threats for the Internet-of-Things. 2019 IEEE 1st international conference on unmanned vehicle systems-Oman (UVS), pp 1–10
14. Teodorovich NN, Stroganova SM, Abramov PS (2017) Methods for detecting and combating small-sized unmanned aerial vehicles., vol 9 (1). Internet magazine "Naukovedenie". http://naukovedenie.ru/PDF/13TVN117.pdf (in Russia)
15. Kaidenko M, Kravchuk S (2019) Creation of communication system for unmanned aerial vehicles using SDR and SOC technologies. 2019 international conference on information and telecommunication technologies and radio electronics (UkrMiCo), pp 1–4, Odessa, Ukraine. https://doi.org/10.1109/UkrMiCo47782.2019.9165422
16. Kaidenko M, Kravchuk S (2021) Autonomous unmanned aerial vehicles communications on the base of software-defined radio. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technology and Systems, vol 152. Lecture Notes in Networks and Systems. Springer, Cham, pp 289–302. https://doi.org/10.1007/978-3-030-58359-0_16
17. European table of frequency allocations and applications in A the frequency range 8.3 kHz to 3000 Ghz approved November 2020. ERC REPORT 25. https://docdb.cept.org/download/2051
18. Calculation of free-space attenuation, recommendation ITU-R P.525–4 (2019). https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.525-4-201908-I!!PDF-E.pdf. Accessed 2016/08
19. Palestini C (2020) Fighting UAVs: finding the magic solution. NATO review, 16 December. https://www.nato.int/docu/review/ru/index.html
20. Kaidenko MM, Roskoshnyi DV (2019) Software Defined radio in communications. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technologies, vol 560. Lecture Notes in Electrical Engineering. Springer, Cham, pp 227–238. https://doi.org/10.1007/978-3-030-16770-7_11
21. Kaydenko N (2013) Adaptive modulation and coding in a broadband wireless access systems. 23th international crimean conference on "Microwave and Telecommunication Technology", on September 8–13, pp 275–276. CriMico'2013, Sevastopol
22. Ilchenko M, Kravchuk S, Kaydenko M (2019) Combined over-the-horizon communication systems. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technologies, vol 560. Lecture Notes in Electrical Engineering. Springer, Cham, pp 121–145. https://doi.org/10.1007/978-3-030-16770-7_6
23. Kravchuk S, Kaidenko M (2016) Features of creation of modem equipment for the new generation compact troposcatter stations, 2016 international conference radio electronics & info communications (UkrMiCo), pp 1–4. IEEE Conference Publications. https://doi.org/10.1109/UkrMiCo.2016.7739634
24. AD9361 RF Agile Transceiver (2021). https://www.analog.com/media/en/technical-documentation/data-sheets/AD9361.pdf. Accessed 2021 June 14

# Estimation of Motion Parameters of Unmanned Aerial Vehicles of Wireless Sensor Networks Based on the Least Squares Method with a Fractional Taylor Series in a "Sliding Window"

**Oleg Tsukanov and Yevgenii Yakornov**

**Abstract** In this paper, to improve the accuracy of estimating the motion parameters of constantly maneuvering quadrocopters as elements of flying wireless sensor networks, a mathematical approach is proposed based on the use of the capabilities of fractional Taylor series, which are still little studied and have not yet found proper practical application in describing complex dynamic processes in various technical applications. At present, for high-precision determination of the coordinates of the motion parameters of unmanned aerial vehicles, with a known motion model, algorithms based on the Kalman-Bucy filter or the least squares method are used, since the potential estimation accuracy of these methods is almost the same. However, in order to estimate the motion parameters of constantly maneuvering aircraft, it is necessary to use the method of estimating from a selection of measurements in a "sliding window". Comparative evaluation using algebraic polynomials, Chebyshev polynomials and fractional polynomials by the method of least squares "in a sliding window" showed that the use of fractional polynomials allows us to evaluate not only changes in coordinates and velocities, but also using fractional derivatives, other parameters that occupy an intermediate position between coordinates, the first and second derivatives with respect to coordinates. The latter makes it possible to improve the accuracy of estimates of the coordinates of the motion parameters of maneuvering unmanned aerial vehicles. Moreover, the most acceptable polynomial for estimation is a polynomial with fractional degrees equal to 2.5. The use of fractional Taylor series in the problem of estimating the motion parameters of constantly maneuvering quadrocopters makes it possible not to use recurrent estimation algorithms with adaptation elements, but to achieve the same goal by changing the degree of the polynomial.

O. Tsukanov (✉) · Y. Yakornov (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Peremohy Avenue 37, Kyiv 03056, Ukraine
e-mail: cukanov-o@ukr.net

Y. Yakornov
e-mail: yakornovits@gmail.com

Source line is missing!!!!
Please update.....

**Keywords** Wireless sensor network · Unmanned aerial vehicles · Motion parameters · Least squares method · Fractional Taylor series · Chebyshev polynomials · Sliding window interval

## 1 Introduction

Wireless flying sensor networks based on miniature unmannedaerial vehicles (UAVs) of the quadrocopter type are increasingly used in telecommunication systems with increased survivability [1]. One of the tasks that developers have to face when creating such networks is to determine the motion parameters (MP) of such UAVs with high accuracy in the presence of destabilizing flight factors such as wind, changes in atmospheric pressure depending on flight altitude, and others. Therefore, for such UAVs, the trajectory of their flight should be considered as the trajectory of an object with a complex type of maneuver and consisting of many sections, each of which must be described by its own system of differential equations, that is, the nature of its movement in space is the trajectory of a constantly maneuvering object.

Obviously, for high-precision determination of coordinates by estimating the MP of a maneuvering UAV, it is necessary to apply adaptive approaches and algorithms. At the same time, the basis of this approach, as shown by the authors in [2, 3], is the recurrent Kalman-Bucy filter.

However, the main disadvantage of the latter is the fact that in the case of a maneuver, the Kalman-Bucy filter will work for some time in the transition section of the estimation, and then in the stationary one. At the same time, with a short-term maneuver, the Kalman-Bucy filter [4, 5] will always work in the transition section. And if the nature of the movement of the UAV is the trajectory of the movement of a constantly maneuvering object, then using algorithms based on the Kalman filter, it is not possible to obtain high-precision estimates of MP [6].

Nevertheless, in our opinion, for high-precision determination of the coordinates of the UAV MP, as elements of wireless sensor networks, algorithms based on the Kalman-Bucy filter or the least squares method (LSM) can be used. The potential estimation accuracy of these methods is almost the same. The optimal approach to assessing MP is to use these methods simultaneously. Moreover, in the following sequence: first, the LSM to obtain an estimate of the measurements, and then the Kalman-Bucy filter to directly estimate the MP. In other words, the Kalman-Bucy filter must be used with a known motion model. But this is only when working on a maneuvering UAV, in the case of a constant maneuver, this approach is not acceptable. Therefore, in order to estimate the AP of constantly maneuvering UAVs, it is necessary to use estimation methods based on a sample of measurements [7].

This approach makes it quite easy to solve the problem of choosing a motion model in the presence of a constant maneuver. At the same time, the issue of ensuring the stability of estimates is the availability of a sample of measurements, and ensuring the required accuracy is the selection of an approximation polynomial. Another important parameter, the value of which determines the estimation error, is the width

of the "sliding window" [8]. This parameter can be set in the form of a time interval or a number of measurements, according to which the approximation coefficients and the estimate of the desired parameter are subsequently determined [9].

The paper considers the problem of choosing a polynomial for estimating MP by LSM. The possibility of using the Puise polynomial or the fractional Taylor series is also being investigated.

## 2 Formulation of the Problem

The problem of choosing the degree of the approximation polynomial depends on the type and intensity of the UAV maneuver [10, 11]. For example, it is believed that if in flight the UAV makes a transition from uniform motion to maneuver, then the degree of the polynomial should be increased. And at the same time, the number of terms of the approximating polynomial increases. Not only the degree of the polynomial changes, but also the values of each coefficient of this polynomial. Note that the degree of the polynomial always changes discretely depending on the estimation error.

Let us set the following task: when estimating the UAV MP: using the Puise series: without increasing the degree of the approximating polynomial, ensure an increase in the accuracy of the estimates [12].

## 3 Main Part

Consider a Puiseux series with one variable. This is a formal algebraic expression of the form:

$$\mathrm{F(x)} = \sum\nolimits_{n=n_o}^{\infty} a_n x^{\frac{n}{m}}. \tag{1}$$

in which the number $n_0$ is an integer, the number m is a natural number (when m = 1 one obtains an ordinary power series), the coefficients a_n are taken from some ring R.

Given the fact that F(x) takes real values, in the case when n < m, the coefficients $a_n$ be complex.

In some practical applications, the variable x cannot be complex, for example, if such a variable is time.

Series (1) or Puiser in the case when the variable x takes a real value, and the value $n_o = 0$ can be represented as the sum of two series, note without a remainder:

$$\mathrm{F(x)} = \mathrm{T(x)} + \mathrm{R(x)} \tag{2}$$

Here T(x) is a Taylor series, R(x) is a series consisting of fractional terms. We call series (2) the fractional Taylor series. The well-known Taylor series in integer derivatives has the form.

$$\mathrm{f}(x) = \mathrm{f}(x_0) + \mathrm{f}(x_0)' (x - x_0) + \mathrm{f}(x_0)'' \frac{(x - x_0)^2}{2!} + ..\mathrm{f}(x_0)^{\mathrm{n}} \frac{(x - x_0)^{\mathrm{n}}}{\mathrm{n}!}. \quad (3)$$

or in general.

$$\mathrm{f}(x) = \sum_{\mathrm{k}=0}^{\mathrm{n}} \mathrm{f}(\mathrm{x}_0)^{\mathrm{k}} \frac{(x - x_0)^{\mathrm{k}}}{\mathrm{k}!},$$

where k = 0, 1, 2, 3, …...

Let us write the representation of the Taylor series (3) for fractional derivatives [4]. For this, the functions f(x) defined on the segment [a,b] can be represented as

$$D^a_{a+}\mathrm{f}(\mathrm{x}) = \frac{1}{\Gamma(1-a)} \frac{d}{dx} \int_a^x \frac{f(x)dt}{(x-t)^a}, \quad (4a)$$

$$D^a_{b-}\mathrm{f}(\mathrm{x}) = \frac{1}{\Gamma(1-a)} \frac{d}{dx} \int_x^b \frac{f(x)dt}{(x-t)^a}, \quad (4b)$$

and are called right (4a) and left (4b) fractional derivatives of order $a$, $0 < a < 1$, respectively left-handed and right-handed. Here $\Gamma(1\text{-}a)$ is Euler's gamma function.

Fractional derivatives in the form (2) are also called [4] Riemann–Liouville derivatives. If there is a general analytical expression for the nth order derivative, the concept of a fractional derivative can be introduced in a natural way by generalizing this expression (when possible) to the case of an arbitrary number.

Then the well-known notation of the first derivative for the functions $\mathrm{f}(\mathrm{x}) = x^k$, we write as

$$\mathrm{f}(x)' = \frac{d}{dx}\mathrm{f}(\mathrm{x}) = \mathrm{k}\, x^{k-1},$$

but in general

$$\frac{d^n}{dx^n} x^k = \frac{k!}{(k-n)!} x^{k-n}, \quad (5)$$

after replacing factorials with gamma functions, can be represented as

$$\frac{d^n}{dx^n} x^k = \frac{\Gamma(k)}{\Gamma(k-n+1)} x^{k-n}. \quad (6)$$

Further, without a strict mathematical justification, we introduce an assumption: if there is a Taylor series for integer derivatives, then a similar series can exist with fractional derivatives or Puiser for $n_o = 0$, and real coefficients.

Then, taking into account the existence of fractional derivatives, such a series has the form:

$$f(x_0) = \sum_{k=0}^{n} f(x_0)^k \frac{\Gamma(k)}{\Gamma(k-n+1)}(x - x_0)^{k-n}, \tag{7}$$

where k = 0, 0.5, 1, 1.5, 2, 2.5…

Let us call the additional terms of the introduced series fractional. Note that for the Taylor series (7) it cannot be stated that the series converges like the series (1). The reason for this is the lack of orthogonality of the terms of the resulting series.

So, for example, for the series (1) of the function $1, x^1, x^2, \ldots . x^n$—are orthogonal, at the same time for series (5), $x^{1/2}$, $x^1$, $\cdots x^{3/2}$, $\ldots$ $x^m$, fractional and integer terms are orthogonal in the other metric. It can be argued that series (7) has extended properties in comparison with series (3).

Note that the authors do not pretend to be mathematically rigorous in the use of series (5). Our task: to find out by modeling that when using the Taylor series (5) to estimate the MP of maneuvering UAVs, the use of the fractional Taylor series to obtain estimates by the least squares will lead to an increase in the accuracy of estimates, compared with using series (1). The hope for this gain is due to the fact that the use of series (5) makes it possible to increase the discreteness of the estimation of the derivatives of the UAV trajectory components.

To begin with, let's compare the fractional Taylor series with his usual series. We also confine ourselves to using series (7) for derivatives not higher than 3/2.

We also leave aside the physical meaning of the coefficients for fractional derivatives and try to explain the essence of the fractional derivative from a mathematical point of view.

For example, for a fractional derivative equal to 1/2, its value takes the average value between the antiderivative and the first derivative. Accordingly, the derivative 1.5 is the average value between the first and second derivatives, etc. At the same time, it can be argued that if, for example, the n-th derivative is equal to 0, and n-1 is non-zero, then the derivative that will occupy an intermediate position will not always be zero.

The preliminary analysis of the physical meaning of the approximation coefficient at a degree of 1.5 carried out by the authors shows that it does not need to be taken into account if the value of this parameter is not used in subsequent calculations, and there is some limitation in the use of fractional Taylor series [13].

Moreover, such cases exist when using LSM with orthogonal Chebyshev polynomials, the attractiveness of which lies in the fact that when increasing the degree of the polynomial, it is necessary to calculate only one coefficient at this degree, and when decreasing, nothing needs to be calculated - just discard the highest term Chebyshev series. At the same time, the coefficients of the series themselves have no

physical meaning, and when using orthogonal Chebyshev polynomials, it is necessary to additionally determine the known physical values: position, velocity, and acceleration [14].

Based on the foregoing, let us analyze the application of the introduced Taylor series (5) with fractional derivatives in relation to the problem of estimating the UAV PM.

## 4 LSM Simulation Results with Fractional and Ordinary Taylor Series

Below are some of the simulation results obtained using the MATLAB system.

Figure 1 shows the result of the approximation of a polynomial of the fourth degree with noise distributed according to the normal law with an error variance equal to 10 $m^2$—green color, on the whole interval by the method of least squares by a polynomial of the second degree - red color,

$$Y = a_0 + a_1 x^1 + a_2 x^2, \tag{8}$$

and by least squares a polynomial of degree 3/2 through fractional degrees 1/2.

$$Y = a_0 + a_1^* x^{0.5} + a_2 x^1 + a_3^* x^{\frac{3}{2}}, \tag{9}$$

where $a_1^*$, $a_3^*$—coefficients of meaningful fractional derivatives—blue color.

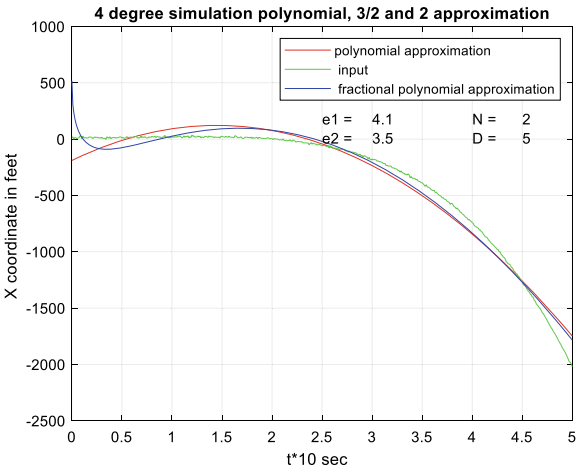



**Fig. 1** Estimation of the UAV trajectory by an algebraic polynomial of the 2nd degree and a fractional polynomial of the degree 3/2 (9) UAV trajectory simulation polynomial of the 4th degree in one coordinate

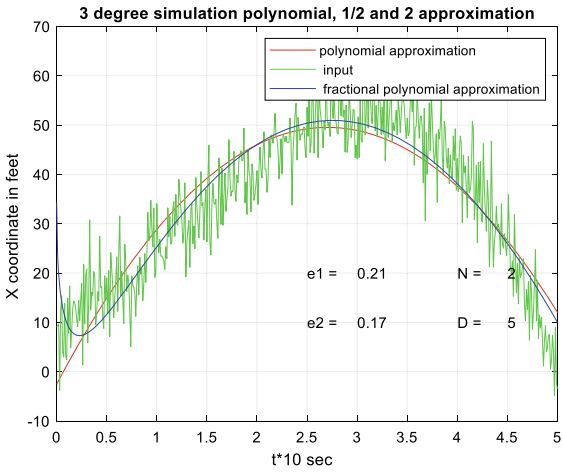



**Fig. 2** Estimation of the UAV trajectory by an algebraic polynomial of the 2nd degree and a fractional polynomial of the degree 1/2. (10) UAV trajectory simulation polynomial of the 3th degree in one coordinate

It can be seen from the graphs that the estimation error by the least squares polynomial (7) decreases by 10–12%.

Figure 2 shows the result of approximating a polynomial of the 3rd degree with noise distributed according to the normal law with an error variance equal to 5 $m^2$—green color, according to the LSM method by polynomial (6)—red color, and a fractional polynomial of the form, blue color.

$$Y = a_0 + a_1 x^{0.5} + a_2 x^1. \tag{10}$$

And in this case, the use of a polynomial with a fractional degree makes it possible to increase the estimation accuracy up to 20%.

Figure 3 shows the results of estimation by a polynomial of the fourth degree with noise distributed according to the normal law with an error variance equal to 5 $m^2$—green color, polynomial (6)—red color and estimation by the least squares polynomial—blue color.

$$Y = a_0 + a_1 x^{0.5} + a_2 x^1 + a_3 x^{\frac{3}{2}} + a_3 x^2. \tag{11}$$

The simulation results show that the use of polynomial (9) instead of (6) makes it possible to increase the estimation accuracy by 20%.

On Fig. 4 shows an estimate by a polynomial of the third degree with noise distributed according to the normal law with an error variance equal to 10 $m^2$ - green color, polynomial (6)—red color, and estimates by LSM by a fractional polynomial.

$$Y = a_0 + a_1 x^{0.5} + a_2 x^1 + a_3 x^{\frac{3}{2}} + a_3 x^2 + a_3 x^{2.5}. \tag{12}$$

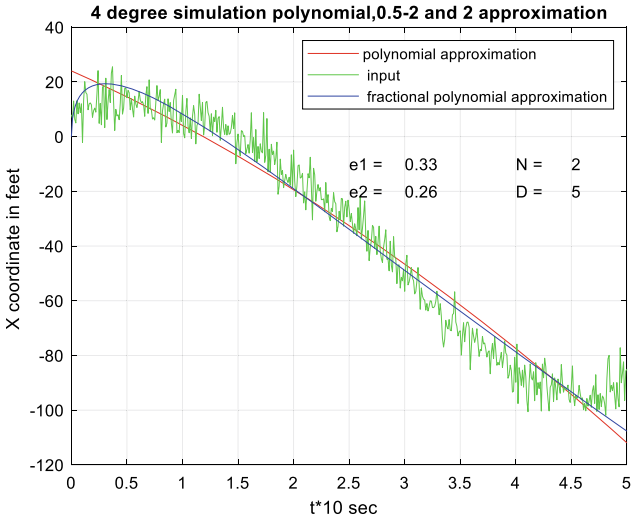



**Fig. 3** Estimation of the UAV trajectory by an algebraic polynomial of the 2nd degree and a fractional polynomial of the degree 2 through 0.5. UAV trajectory simulation polynomial of the 4th degree in one coordinate

In this case, the estimation accuracy increases by 12%.

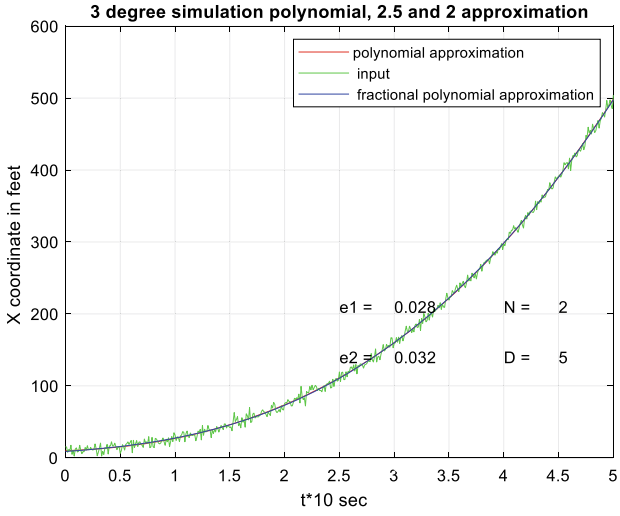



**Fig. 4** Estimation of the UAV trajectory by an algebraic polynomial of the 2nd degree and a fractional polynomial of the degree 2.5 (12) UAV trajectory simulation polynomial of the 3th degree in one coordinate

Thus, the use of polynomials (9)–(11) makes it possible to increase the estimation accuracy, and the use of polynomial (12) is not justified for a simulation trajectory in the form of an algebraic polynomial of the 3rd degree.

Figure 4 shows the estimate by a polynomial of the third degree with noise distributed according to the normal law with a root-mean-square error equal to 10 $m^2$—green color, a polynomial also of the third degree—red color, and estimates by the least squares by a fractional polynomial of degree 2.5.

Figure 5 shows an estimate of a polynomial of degree 3 with noise distributed according to the normal law with a mean square error equal to 10 $m^2$—green color, a polynomial of also degree 3—red color, and estimates of the least squares with a fractional polynomial of degree 2.5.

In this case, the coincidence of the degrees of the simulation polynomial and the approximating polynomial, the accuracy of estimates by the fractional polynomial is worse by 15%, the result is expected.

For clarity, we summarize the simulation results in Table 1.

It can be seen from the table that the use of the proposed fractional Taylor series in the problem of estimating the motion parameters of constantly maneuvering UAVs, when their motion trajectory is described by a polynomial of a higher degree, makes it possible to improve the accuracy of the estimates. Moreover, for each class of UAVs, it is possible to fix the degree of the polynomial for estimating the MP. This will greatly simplify the structure of the software without sacrificing accuracy.

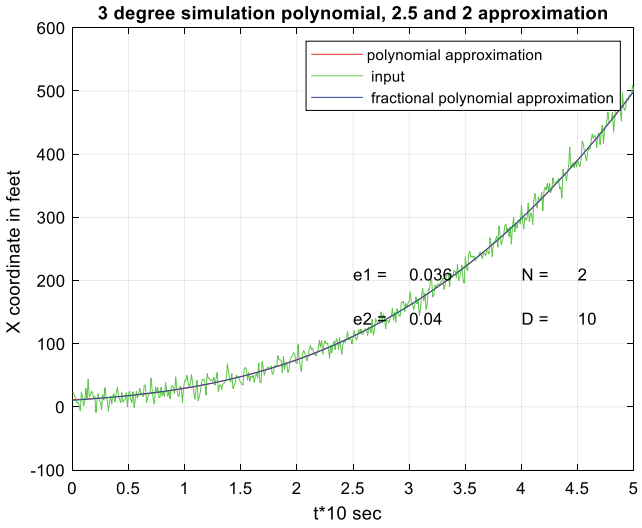



**Fig. 5** Estimation of the UAV trajectory by an algebraic polynomial of the 2nd degree and a fractional polynomial of the degree 2.5 (12) UAV trajectory simulation polynomial of the 3th degree in one coordinate

**Table 1** Simulation results

| Specification | Values | | | | |
|---|---|---|---|---|---|
| The degree of approximation of the simulation polynomial | 4 | 3 | 3 | 3 | 3 |
| The degree of approximation of a polynomial with integer degrees | 2 | 2 | 2 | 2 | 2 |
| The degree of approximation of a polynomial with fractional degrees | 3/2 | 1/2 | 2 | 2.5 | 2.5 |
| Error of approximation by a polynomial with integer powers | 4.1 | 0.21 | 0.33 | 0.028 | 0.036 |
| Error approximation by polynomial with fractional powers | 3.5 | 0.17 | 0.26 | 0.032 | 0.04 |
| Improving accuracy, % | 15% | 20% | 22% | −13 | −11% |

Note that the least squares estimation using fractional Taylor series is preferable when the order of the simulation polynomial is higher than the approximation polynomial. In the case when the degree of the imitating polynomial and the approximation polynomial coincide, then the use of fractional Taylor series does not make sense. But it is here that the main advantage of using the fractional Taylor series manifests itself, which lies in the fact that in the general case it is not necessary to know the order of the approximation polynomial. In other words, if there are requirements for the accuracy of the resulting estimates, then the degree of the fractional Taylor series should be chosen in advance, which will ensure an increase in accuracy compared to the known Taylor series.

## 5 Results of Modeling by Chebyshev Polynomials Fractional and Ordinary Taylor Series

Let us compare the accuracy of estimation by LSM now in the "sliding window" by Chebyshev polynomials, fractional and ordinary Taylor series. For comparison, consider Chebyshev polynomials of the first kind of the second and third, polynomials based on the fractional Taylor series and polynomials based on the traditional Taylor series or simply algebraic.

The simulation uses a polynomial based on a Taylor series of the third degree with noise distributed according to the normal law and an error variance equal to 30 $m^2$.

Algebraic polynomials of the second and third order are chosen as approximation polynomials:

$$\mathrm{y(t)} = a_0 + a_1 t + a_2 t^2,$$

$$\mathrm{y(t)} = a_0 + a_1 t + a_2 t^2, + a_1 t^3. \tag{13}$$

Chebyshev polynomials of the first kind of the second and third order

$$T_o(x) = T_1(t) = x;\ T_2(x) = 2x^2 - 1;\ T_3(x) = 4t^3 - 3x,$$

$$T_n(x) = 2x\,T_n(x) - T_{n-1}(x), \tag{14}$$

$$y(t) = a_o + a_1 t + a_2(2t^2 - 1),$$

$$y(t) = a_o + a_1 t + a_2(2t^2 - 1) + a_3(4t^3 - 3x). \tag{15}$$

Fractional polynomials based on the Taylor series

$$y(t) = a_0 + a_1 t^{0.5} + a_2 t^{1.5}, \tag{16}$$

$$y(t) = a_0 + a_1 t^{0.5} + a_2 t + a_3 t^{1.5} \tag{17}$$

It is assumed that the UAV constantly maneuvers in accordance with the 3rd order polynomial and the parameters of noise distributed according to the normal distribution law with zero mean and variance equal to 30 $m^2$.

It is known that Chebyshev polynomials exist in the interval [–1; 1].

LSM estimation with any polynomials requires matrix inversion. For Chebyshev polynomials on this interval, with a large width of the sliding window, difficulties arise with matrix inversion. In this case, the Seidel method should be used. It is also known that Chebyshev polynomials have the best approximation among all known polynomials.

In the case of using polynomials (16, 17), there are also problems with matrix inversion. We recommend bringing the estimation interval to the values [1, 5] and then there are no problems with matrix inversion.

Figure 6 shows the results of estimating the trajectory of motion by an algebraic polynomial—red color, a Chebyshev polynomial—blue color, an algebraic polynomial—green color. Degree of fractional polynomial Nfr = 1/2, polynomial (15), degree of algebraic polynomial Nal = 3, degree of Chebyshev polynomial Nch = 3. E cheb is the error of estimates by the Chebyshev polynomial, E frac is the error of estimates by the fractional Taylor series, E alg are the errors of estimates by the algebraic polynomial, W = 12 is the "sliding window" interval.

The simulation results show that the accuracy of fractional polynomial estimates is 13% higher than that of the algebraic polynomial and 12% higher than that of the Chebyshev polynomial, the latter being due to the width of the "sliding window" interval. At the same time, the accuracies of the estimates by the fractional and polynomial Chebyshev coincide.

Figure 7 shows the results of estimates by an algebraic polynomial—red color, Chebyshev polynomial—blue color, algebraic polynomial—green color.

Fractional polynomial (14) has degree Nfr = 3/2, degree of algebraic polynomial Nal = 3, degree of Chebyshev polynomial Nch = 3, Echeb—estimation error by

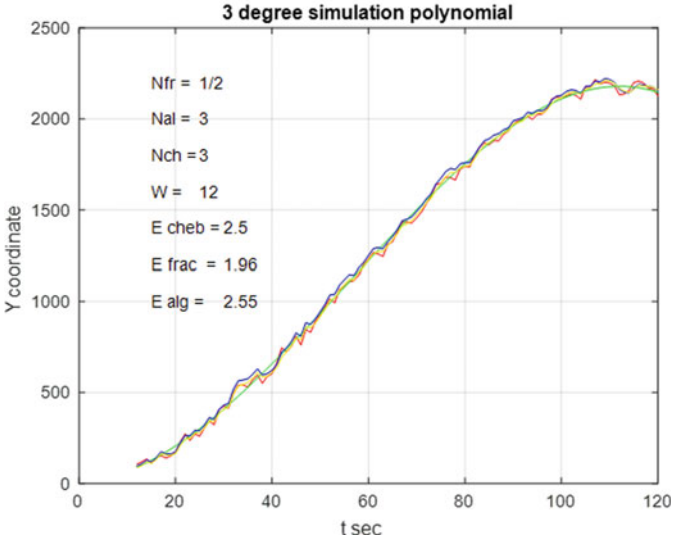



**Fig. 6** Estimates of the UAV trajectory by an algebraic degree polynomial, a Chebyshev polynomial, and a fractional degree polynomial (16) in a sliding window for 12 measurements

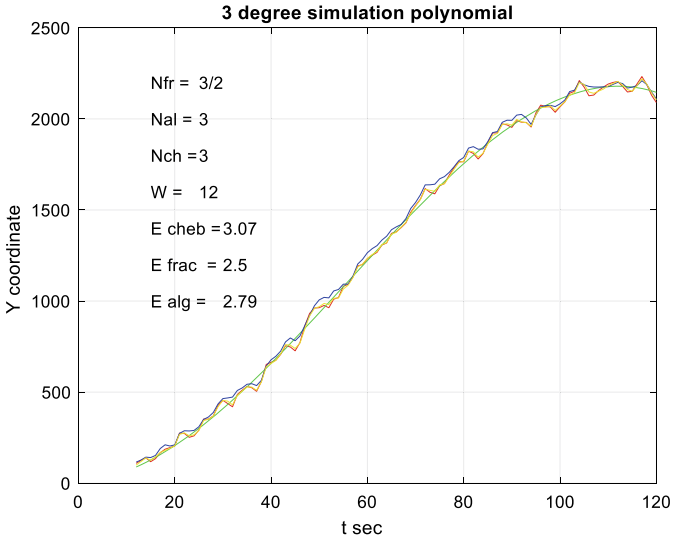



**Fig. 7** Estimates of the UAV trajectory by an algebraic degree polynomial, a Chebyshev polynomial, and a fractional degree polynomial (17) in a sliding window for 12 measurements

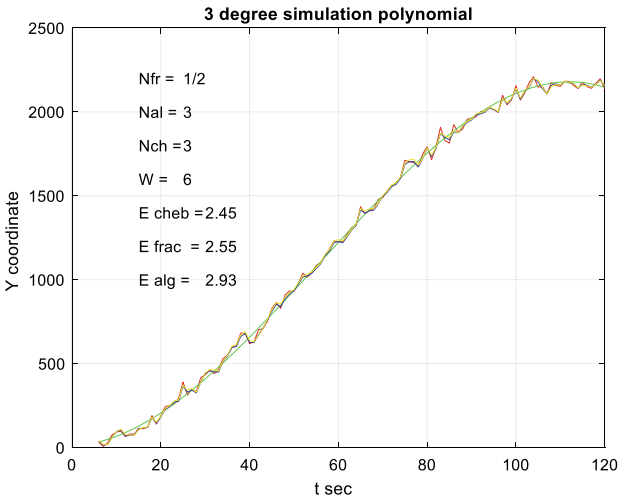



**Fig. 8** Estimates of the UAV trajectory by an algebraic degree polynomial, a Chebyshev polynomial, and a fractional degree polynomial (16) in a sliding window for 6 measurements

Chebyshev polynomial, Efrac—estimation error by fractional Taylor series, Ealg—estimation errors by algebraic polynomial, W is the "sliding window" interval.

The accuracy of estimates by the fractional polynomial is 19% higher than the algebraic one and 11% higher than the Chebyshev polynomial, the latter is due to the width of the "sliding window" interval. The maximum estimation accuracy is achieved for the polynomial (4), which is due to the "sliding window" interval W = 12.

Figure 8. shows the results of estimation by the algebraic polynomial—red color, by the Chebyshev polynomial—blue color, by the algebraic polynomial—green color. Here the degree of the fractional polynomial is Nfr = 1/2, (13), the degree of the algebraic polynomial is Nal = 3, the degree of the Chebyshev polynomial is Nch = 3/2.

The maximum accuracy of estimates is achieved by using a fractional Taylor polynomial of order 1/2.

Here we should dwell on the polynomial (13) in which all degrees are fractional and are multiples of 1/2. The use of this polynomial, under certain conditions, provides the maximum accuracy of estimates in comparison with the Chebyshev polynomials and the polynomial based on the Taylor series.

The estimation accuracy of the Chebyshev polynomial is 4% higher than that of the fractional polynomial and 5% higher than the algebraic polynomial.

Figure 9 shows the results of estimation by the algebraic polynomial - red color, by the Chebyshev polynomial - blue color, by the algebraic polynomial - green color.

Fractional polynomial (17) of degree Nfr = 1/2, degree of algebraic polynomial Nal = 3, degree of Chebyshev polynomial Nch = 2, Echeb—estimation error by

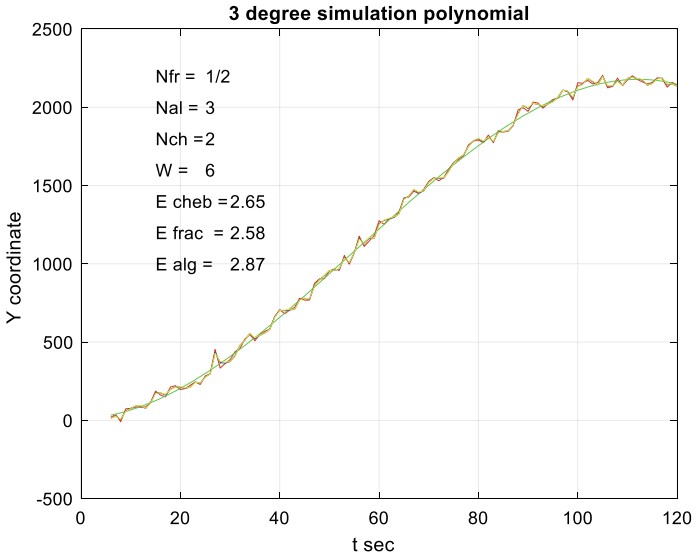



**Fig. 9** Estimates of the UAV trajectory by an algebraic degree polynomial, a Chebyshev polynomial degree 2, and a fractional degree polynomial (16) in a sliding window for 6 measurements

Chebyshev polynomial, Efrac—estimation error by fractional Taylor series, Ealg—estimation errors by algebraic polynomial, W = 6—width of the "sliding window" interval. The maximum accuracy is achieved by a fractional polynomial of order ½ and the Chybyshev polynomial, while the accuracy of estimation by an algebraic polynomial is 5% less.

Figure 10 shows the results of estimation by the algebraic polynomial - red color, by the Chebyshev polynomial - blue color, by the algebraic polynomial - green color.

Fractional polynomial (17) of degree Nfr = 3/2, degree of algebraic polynomial Nal = 3, degree of Chebyshev polynomial Nch = 2, Echeb—estimation error by Chebyshev polynomial, Efrac—estimation error by fractional Taylor series, Ealg—estimation errors by algebraic polynomial, W = 6—width of the "sliding window" interval. The maximum accuracy is achieved by a fractional polynomial of order ½ and the Chybyshev polynomial, while the accuracy of estimation by an algebraic polynomial is 8% less.

And finally, Fig. 11. shows the estimates by the algebraic polynomial - red color, the Chebyshev polynomial - blue color, the algebraic polynomial - green color. Fractional polynomial (14) of degree Nfr = 3/2; degree of algebraic polynomial Nal = 3; is the width of the "sliding window" interval.

The accuracy of estimates by the Chebyshev polynomial is 7% higher than the fractional one and 11% higher than the algebraic polynomial.

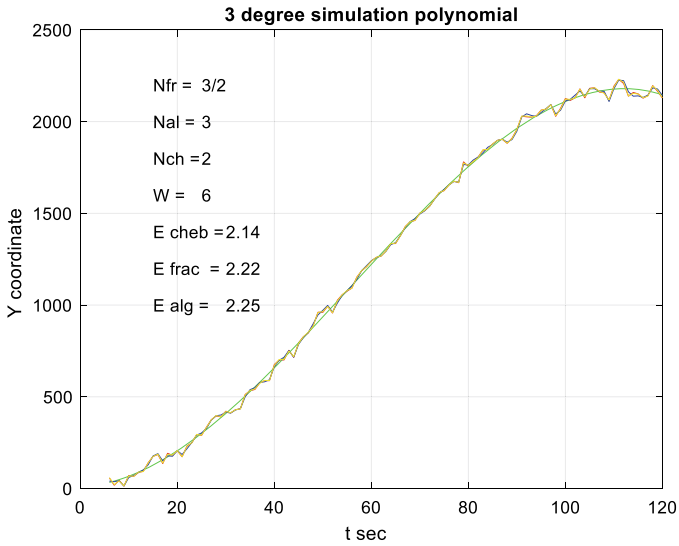



**Fig. 10** Estimates of the UAV trajectory by an algebraic degree polynomial, a Chebyshev polynomial degree 2, and a fractional degree polynomial (17) in a sliding window for 6 measurements

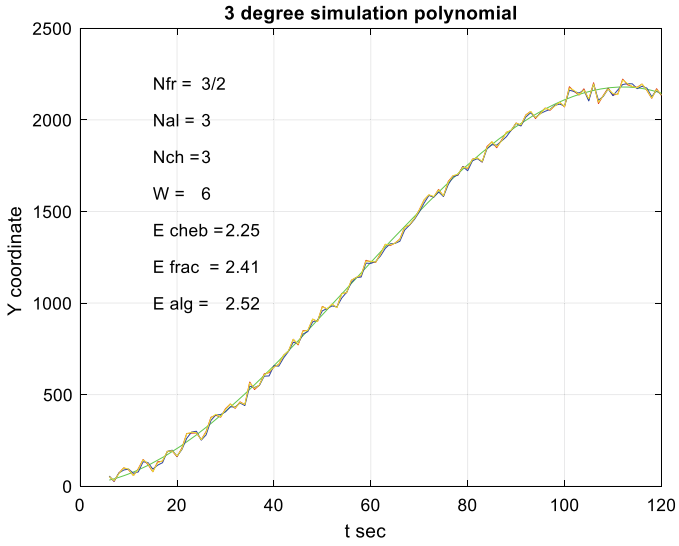



**Fig. 11** Estimates of the UAV trajectory by an algebraic degree polynomial, a Chebyshev polynomial degree 2, and a fractional degree polynomial (17) in a sliding window for 6 measurements

## 6 Conclusion

1. The results of the research have shown that the use of fractional Taylor series can significantly expand the class of problems for using LSM in the problems of estimating the MP of quadrocopters as elements of wireless sensor networks.
2. The use of fractional Taylor series in the problem of estimating the MP of maneuvering quadrocopters as elements of wireless sensor networks makes it possible not to use recurrent estimation algorithms with adaptation elements by changing the degree of the polynomial.
3. It follows from the simulation results that for a constantly maneuvering UAV that constantly maneuvers, the most acceptable estimation polynomial is the LSM with fractional powers of 2.5.
4. Further research is needed on fractional polynomials exclusively with fractional degrees above degree 1.5.
5. The use of fractional polynomials for estimating in practice is limited by the lack of physical meaning of fractional derivatives.
6. Nevertheless, the use of fractional polynomials makes it possible to obtain a greater estimation accuracy compared to Chebyshev polynomials with a large width of the "sliding window".
7. Fractional polynomials have broad prospects for solving the problem of estimating the motion parameters of quadrocopters.

## References

1. Romaniuk V, Lysenko O, Romaniuk A, Zhuk O (2020) Increasing the efficiency of data gathering in clustered wireless sensor networks using UAV. Inf Telecommun Sci 11(1):102–107. https://doi.org/10.20535/2411-2976.12020.102-107
2. Yakornov Y, Tsukanov O (2019) Sustainable algorithm for estimating the motion parameters of unmanned aerial vehicles. International conference on information and telecommunication technologies and radio electronics (UkrMiCo) – proceedings, September, pp 1–5. https://doi.org/10.1109/UkrMiCo47782.2019.9165389
3. Kilbas A, Srivastava H, Trujillo J (2006) Theory and applications of fractional differential equations. Elsevier, Amsterdam. https://doi.org/10.1109/APUAVD.2015.7346621
4. Legendre AM (1806) Nouvelles méthodes pour la dé termination des orbites des comètes. Sur la methode des moindres carrés, pp. 72–80. Appendice, Paris. (in France)
5. Gauss KF (1958) Izbrannye geodezicheskie sochineniya: T.1. GV Bagratuni (ed) Selected geodetic works, vol. 1. Geodesisdat, Moscow. (in Russian)
6. Adrian R (1908) Research concerning the probabilities of the errors which hap pen in making observations. The analyst or mathematical museum 1(4):93–109
7. Idelson NI (1947) Sposob naimen'shikh kvadratov i teoriya matematicheskoy obrabotki nablyudeniy (Method of least squares and the theory of ma thematical treatment of observations). Geodesisdat, Moscow (in Russian)
8. Elyasberg PE (1976) Opredelenie dvizheniya po rezul' tatam izmereniy (The definition of motion by the results of measurements). Nauka, Main Editorial Board of Physical and Mathematical Literature, Moscow (in Russian)

9. Padve B (2013) Elementy teorii veroyatnostey i matematicheskoy statistiki (Elements of probability theory and mathematical statistics). SSGA, Novosibirsk (in Russian)
10. Tovkach O, Zhuk S, Neuimin O (2020) Estimation of UAV movement parameters based on TDOA measurements of the sensor network in: the presence of Ab normal measurements. 2020 IEEE 15th international conference on advanced trends in radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, pp 235–239. https://doi.org/10.1590/jatm.v13.1242
11. Tovkach I, Zhuk S (2020) Adaptive filtration of the UAV movement parameters based on the AOA-measurement sensor networks. Int J Aviat Aeronaut Aerosp. https://doi.org/10.15394/ijaaa.2020.1497
12. Masieroa A, Fissorea F, Guarnieria A, Pirottia F, Vettorea A (2015) UAV positioning and collision avoidance based on RSS measurements, The inter national archives of the photogrammetry, remote sensing and spatial information sciences, vol XL-1/W4, pp 219–225
13. Sierociuk D (2006) Fractional Kalman filter algorithm for states, parameters and order of fractional system estimation. Int J Appl Math Comput Sci 16(1):129–140. http://eudml.org/doc/207770
14. Lysenko O, Novikov V, Alekseeva I (2015) Optimization of unmanned aerial vehicle path for wireless sensor network data gathering. 3rd international conference: actual problems of unmanned aerial vehicles developments (APUAVD-2015), Proceedings, pp 280–283

# Modern Challenges in Radio Electronics Technologies

# Mixed Coupling in Trisection and Quadruplet Bandpass Filters

**Alexander Zakharov and Michael Ilchenko**

**Abstract** This article solved the inverse problem for trisection and quadruple bandpass filters (BPFs) with mixed cross coupling. It allows for a given placement of transmission zeros and known main coupling coefficients to determine mixed cross coupling $K = K_m + K_e$, containing magnetic and electrical components. Based on the obtained solution, it was established that trisection BPF has ten different options for placing two transmission zeros on the complex plane $S = \sigma + j\Omega$. It is shown that the considered trisection and quadruplet BPFs can have a second-order transmission zero on the $j\Omega$ axis, which provides a deeper attenuation pole at insertion loss curve. With the help of the obtained inverse problem solution, some restrictions are established on the possible options for the placement of three transmission zeros of quadruplet BPF with mixed cross coupling $K_{14}$: transmission zeros cannot be placed on the $\sigma$ axis; two of the three transmission zeros on the $j\Omega$ axis cannot be equidistantly relative to $S = 0$. It is found that to obtain a flat group delay, the transmission zeros mast be located in the $S$ plane at the corners of an isosceles triangle, the vertex of which lies on the $j\Omega$ axis, and the sides intersect the $\sigma$ axis. In this case, in addition to the flat group delay, the insertion loss curve has an attenuation pole. Theoretical results are validated with two microstrip quasi-inline trisection BPFs and one stripline quadruplet BPF.

**Keywords** Frequency responses · Inverse problem · Mixed coupling · Quadruplet filter · Transmission zeros · Trisection filter

A. Zakharov (✉) · M. Ilchenko (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: azakharov217@gmail.com

M. Ilchenko
e-mail: ilch@kpi.ua

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_22

## 1 Introduction

Bandpass filters are one of the main elements of communication systems. Recently, they have received considerable attention in the scientific literature [1–5]. Small-sized filters like filters designed on coaxial dielectric resonators [6], monoblock ceramic filters [7], stripline filters [8], surface acoustic wave (SAW) filters [9], multilayer ceramic filters [10, 11], and microstrip filters [12–15] are in the great demand.

The papers of R.M. Kurzrok about trisection BPF [16] and quadruplet BPF [17] are great importance for filter theory and practical application. In these papers, formulas are established that express the analytical dependences between coupling coefficients and transmission zeros on the complex plane $S = \sigma + j\Omega$. Trisection BPF has one transmission zero on the $j\Omega$ axis. The $K_{14}$ cross-coupled quadruplet BPF has two transmission zeros that are equidistantly on the $\sigma$ or $j\Omega$ axis. These formulas allow solving the inverse problem, i.e. determine the cross-coupling coefficient for the given transmission zeros and main coupling coefficients. The properties established in [16, 17] are repeated in CT and CQ filters obtained by cascading trisection [18] or quadruplet [19] BPFs.

The use of mixed couplings $K = K_m + K_e$ in BPFs allows us to reduce the number of resonators $N$ required to obtain a given number of transmission zeros [20–37]. Such couplings are also called "resonant" [20, 30] or "frequency-dependent" [22–27]. The properties of BPFs with mixed couplings were first described in [20, 21]. In a later work [22], a linear representation form of the mixed coupling $m(\Omega) = m_0 - a\Omega$ was proposed, which made it possible to introduce it into the coupling matrix and calculate the frequency responses based on this matrix. In [22], mixed couplings between adjacent resonators were considered.

In a series of articles [23–26], a linear representation form of mixed coupling is used, but for filters with cross couplings. In these articles, parametric synthesis is used, which is also called optimization [38]. In parametric synthesis, the coupling structure of the filter is given, the elements of the coupling matrix are variable. These elements are changed according to a certain algorithm, which leads to a given location of the transmission zeros. This synthesis method has considerable generality, which has allowed many important placements of transmission zeros to be established for BPFs with different coupling structures and $N$ resonators.

However, the parametric synthesis method is not analytical. It leads to a positive result only if the given arrangement of the transmission zeros is physically feasible. This method does not provide answers to a number of important questions related to the possible placement of transmission zeros. Can trisection and quadruplet BPFs transmission zeros be second-order? Can the transmission zeros (one or more) of the quadruplet BPF with mixed cross coupling $K_{14}$ be located on the $\sigma$ axis? Can a quadruplet BPF with all mixed couplings (without I/O coupling), containing three cross couplings, have four transmission zeros? The location of the transmission zeros of quadruplet BPF with mixed coupling $K_{14}$ and flat group delay [25] has not yet been established.

Apparently, the first analytical method for the synthesis of BPFs with mixed couplings was proposed in [27]. But it is applicable to filters with a special coupling structure, and in some cases leads to BPF with an excessive number of resonators. So, for diplexer the fifth- and sixth-order BPFs were synthesized, which have two one-sided transmission zeros. For such an arrangement of transmission zeros, it is sufficient to use BPF with $N = 3$ [24, 30].

In [28] the issue of cascading connection of fundamental building blocks was considered, which include trisection and quadruplet BPF with mixed cross coupling. However, the properties of these filters have not been fully studied. So, for trisection BPF, the expression for the element $Y_{12}$ of admittance matrix [$Y_{11}$, $Y_{12}$, $Y_{21}$, $Y_{22}$] is given. A detailed analysis of the equation $Y_{12} = 0$, defining the transmission zeros of this filter, was not carried out, and the inverse transition from transmission zeros to coupling coefficients was not considered. This led to some inaccuracy. The quadruplet BPF with mixed cross-coupling analysis was practically not carried out. There are no formulas expressing the relationship between coupling coefficients and transmission zeros. As a result, six resonators ($N = 6$) are used to realize a BPF with flat group delay and one pole of attenuation at insertion loss curve. Note that to implement such frequency responses, it is sufficient to use four resonators [25].

It can be stated that the inverse problem of transition from transmission zeros to coupling coefficients of trisection and quadruplet BPF with mixed cross coupling has not been solved at present in analytical form. This problem is the subject of this chapter, which is organized as follows. Section 2 discusses the general relationships that define the transmission zeros of BPFs with mixed couplings, and also determines the slope parameter $a$ of the mixed coupling in a linear representation form. Section 3 solves the inverse problem for trisection BPF with mixed cross coupling. Section 4 solves the inverse problem for quadruplet BPF with mixed cross coupling $K_{14}$. The conclusion is given in Section 5.

## 2 Direct and Inverse Problems

$N$-order BPF is characterized by the coupling coefficients $K_{ij}$ ($i, j = 1, 2, \ldots, N$), the external quality factor of the end resonators $Q_{e1}$, $Q_{eN}$, the central frequency of the passband $f_0$, the passband width $BW$, and the fractional bandwidth $FBW = BW/f_0$.

### *2.1 General Relations*

The normalized coupling coefficients $m_{ij} = K_{ij}/FBW$, the external Q-factors $q_{e1} = Q_{e1}FBW$, $q_{eN} = Q_{eN}FBW$, and the complex frequency $S = \sigma + j\Omega$, where $\Omega = (f/f_0 - f_0/f)/FBW$ is the normalized frequency, are used for BPF description.

If the impedance $(N \times N)$ matrix $\left[\tilde{Z}\right]$ is known, then the expression for the transmission function $S_{21}$ of the filter can be represented in the matrix form [39]

$$S_{21} = 2[\tilde{Z}]_{N1}^{-1} \Big/ \sqrt{q_{e1} q_{en}}, \tag{1}$$

where $\left[\tilde{Z}\right]_{N1}^{-1}$ is the minor of the element located at the intersection of the $N$th row and the first column of the inverse matrix $\left[\tilde{Z}\right]^{-1}$. Computing the inverse matrix is a rather time-consuming procedure [40]. Considering that the impedance matrix $\left[\tilde{Z}\right]$ and the admittance matrix $\left[\tilde{Y}\right]$ are mutually reversible [41]: $\left[\tilde{Y}\right] = \left[\tilde{Z}\right]^{-1}$, we can simplify (1):

$$S_{21} = 2M_{N1} \Big/ \sqrt{q_{e1}\, q_{en}}, \tag{2}$$

where $M_{N1}$ is the minor of the admittance matrix $\left[\tilde{Y}\right]$, which can be represented in the form [39]

$$[\tilde{Y}] = \begin{bmatrix} 1/q_1 + S & -jm_{12} & \dots & -jm_{1N} \\ -jm_{21} & S & \dots & -jm_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ -jm_{N1} & -jm_{N2} & \cdots & 1/q_1 + S \end{bmatrix}. \tag{3}$$

To determine the transmission zeros of BPF by using the admittance matrix (3), we have to equate to zero its minor $M_{N1}$, which is the determinant of a square matrix of size $(N-1) \times (N-1)$

$$M_{N1} = \begin{vmatrix} m_{12} & m_{13} & \cdots & m_{1N} \\ m_{22} & S' & \cdots & m_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ m_{(N-1)2} & m_{(N-1)3} & \cdots & m_{(N-1)N} \end{vmatrix} = 0, \tag{4}$$

where $S' = S/j$ and $S'^2 = -S^2$. Note that the impedance matrix $\left[\tilde{Z}\right]$ (1) was arbitrarily chosen as the initial one, and then the transition to the admittance matrix $\left[\tilde{Y}\right]$ was made. The result would be the same if we first select the matrix $\left[\tilde{Y}\right]$ (3), and then go to the matrix $\left[\tilde{Z}\right]$. In both cases the transmission zeros are defined by expression (4).

Let us consider examples of application (4) in some BPFs with simple cross coupling. For the trisection BPF, the minor (4) has the form

$$M_{31} = \begin{vmatrix} m_{12} & m_{13} \\ S' & m_{23} \end{vmatrix} = m_{12}\, m_{23} - m_{13} S'. \tag{5}$$

Equating minor $M_{31}$ (5) to zero, we obtain the expression for the transmission zero $S_1' = m_{12}m_{23}/m_{13}$ or $\Omega_1 = m_{12}m_{23}/m_{13}$, which represents the known formula of R. Kurzrok [16]. If the product of the signs of the coupling coefficients $m_{12,}$ $m_{23,}$ $m_{13}$, corresponds to "+", then the transmission zero on the $j\Omega$ axis is right-hand, $\Omega > 0$. If the specified product of signs corresponds to "–", then the transmission zero on the $j\Omega$ axis is left-hand.

The minor (4) for the quadruplet BPF with one cross coupling $m_{14}$ has the form

$$M_{41} = \begin{vmatrix} m_{12} & m_{13} & m_{14} \\ S' & m_{23} & m_{24} \\ m_{32} & S' & m_{34} \end{vmatrix} = m_{14} S'^2 + m_{12}\, m_{23} m_{34} - m_{14} m_{23} m_{32}. \tag{6}$$

Equating (6) to zero, and taking into account that $S'^2 = -S^2$, we successively obtain the second formula of R. Kurzrok [17]: $S^2 = (m_{12}m_{23}m_{34}/m_{14}) - m_{23}^2$, $S_{12} = \pm\,[(m_{12}m_{23}m_{34}/m_{14}) - m_{23}^2]^{1/2}$. The reliability (4) is confirmed.

Since the condition $|m_{12}m_{23}m_{34}/m_{14}| > m_{23}^2$ holds, then the sign of the radical expression is determined by the sign of the first term. If the product of signs of all coupling coefficients corresponds to "+", then the transmission zeros are real numbers $S_{12} = \sigma_{12}$ ($\sigma_1 = -\sigma_2$) and this filter is a delay line filter. If the indicated product of signs corresponds to "–", then two transmission zeros are located at the real frequencies $S_{12} = j\Omega_{12}$ ($j\Omega_1 = -j\Omega_2$), and this filter is a quasi-elliptic filter. To determine the location of transmission zeros of BPFs with simple couplings, the phase analysis method is often used [42–44].

## 2.2 *Mixed Coupling Coefficient*

The mixed coupling coefficient $K$ [39] is the sum of magnetic $K_m$ and electrical $K_e$ components:

$$K = K_m + K_e = K_m - |K_e|. \tag{7}$$

The magnetic component $K_m$ is assigned a "+" sign, and the electrical component $K_e$ is assigned a "–" sign.

The linear presentation form of mixed coupling was proposed in [22]

$$m(\Omega) = m_0 - a\Omega, \tag{8}$$

where $m_0 = (K_m + K_e)/FBW$, $a$ characterizes the slope of the straight line (8) in the $(m, \Omega)$ plane.

The linear presentation (8) allows us to introduce mixed coupling in impedance, admittance, coupling matrixes and calculate frequency responses. It is assumed that the slope of the straight line (8) can be both positive and negative.

Currently, parameter $a$ is not expressed in terms of the mixed coupling components $K_m$ and $K_e$. To establish this dependence, we use the pattern of the transmission zero location $f_z$ generated by the mixed coupling in two order BPF [31, 32]

$$f_z = f_0\sqrt{K_m / |K_e|}. \tag{9}$$

Using Eq. (9), we express the transmission zero position on the $\Omega$ axis

$$\begin{aligned}\Omega_z &= \left(f_z / f_0 - f_0 / f_z\right) / FBW \\ &= \left(\sqrt{K_m / |K_e|} - 1 \Big/ \sqrt{K_m / |K_e|}\right) \Big/ FBW = m_0 \Big/ \sqrt{K_m / |K_e|}.\end{aligned}$$

Through two points on the plane $(m, \Omega)$, one with coordinates $(\pm m_0, 0)$, and the other with coordinate $(0, \pm \Omega_z)$, we can draw a straight line (8). The coefficient $a$ of this line is equal to the ratio:

$$a = m_0 \Big/ \Omega_z = \sqrt{K_m |K_e|}. \tag{10}$$

The parameter $a$ (10) is always positive, and the slope of line (8) is negative.

To provide a mixed coupling between resonators, various lumped circuits with elements $L$ and $C$ can be used. In all cases, the input susceptance of these circuits has a positive slope, and the coupling coefficient between the resonators has a negative slope. If $K_m$ and $K_e$ components of the mixed coupling mutually compensate each other ($K_m = |K_e|$) and $K = 0$ ($m_0 = 0$), then we get from (10)

$$a = K_m = |K_e| \tag{11}$$

and straight line Eq. (8) takes the form:

$$m(\Omega) = -K_m\Omega = -|K_e|\Omega. \tag{12}$$

The reverse transition from the mixed coupling form (8) to (7) is carried out as follows:

$$K = m_0 FBW; \tag{13a}$$

$$K_m,\ K_e = K \Big/ 2 \pm \sqrt{(K/2)^2 + a^2}. \tag{13b}$$

The "+" sign in (13b) corresponds to $K_m$, and the "–" sign to $K_e$.

If all couplings in $N$-order BPF are mixed, then it is necessary to replace all $m_{ik}$ with the linear form of the mixed coupling (8) in determinant (4): $m_{0ik} - a_{ik}S'$. In this case, the equation $M_{N1} = 0$ (4) will represent an algebraic equation of degree $(N - 1)$

$$S'^{N-1} + A_2 S'^{N-2} + \ldots + A_{N-1}S' + A_N = 0 \tag{14}$$

where the polynomial coefficients $A_i$ are real numbers depending on BPF coupling coefficients. Equation (14) defines $(N - 1)$ transmission zeros. For BPF with simple couplings, the equation $M_{N1} = 0$ is an algebraic equation of degree $(N - 2)$ that defines a smaller number $(N - 2)$ of transmission zeros. The problem of determining the transmission zeros of $N$-order BPFs that contain simple and mixed couplings is to determine the roots of the algebraic Eq. (14) of degree $(N - 1)$.

In order to have the predicted transmission zeros $S_1\prime, S_2\prime, \ldots, S'_{N-1}$ for the discussed BPFs, we should solve the inverse problem. For this, we use the equivalent representation form of the polynomial (14)

$$S'^{N-1} + A_2 S'^{N-2} + \ldots + A_{N-1}S' + A_N = (S' - S'_1)(S' - S'_2)\ldots(S' - S'_{N-1}). \tag{15}$$

The relationship between the zeros of $S'_1$ and the coefficients of the polynomial (15) is expressed by the Vieta's formulas [40]

$$\begin{aligned} A_2 &= -\left(S'_1 + S'_2 + \ldots + S_{N-1}\right) \\ A_3 &= S'_1 S'_2 + S'_1 S'_3 + \ldots + S'_{N-2} S'_{N-1} \\ &- - - - - - - - - - - - - - - - - - \\ A_N &= (-1)^{N-1} S'_1 S'_2 S'_{N-1} \end{aligned} \tag{16}$$

Since the polynomial coefficients $A_i$ are determined using (16), it is possible to find the coupling coefficients, which represents the inverse problem solution. Inverse problem (16) is applicable to BPFs of different orders with different coupling coefficients (simple and mixed). We use these equations for third- and fourth-order BPFs, which only have one mixed cross coupling.

## 3 Transmission Zeros of Trisection BPF with Mixed Cross-Coupling

If the mixed cross-coupling is used in the trisection BPF (Fig. 1a), then it is necessary to replace $m_{13}$ with $m_{13}(\Omega) = m_{013} - a_{13}S''$ in (5), which leads to the algebraic equation of the second degree

$$S'^2 - (m_{013}/a_{13})S' + m_{12}m_{23}/a_{13} = 0\,. \tag{17}$$

The solution of (17) is an analysis problem that determines two transmission zeros $S'_1$ and $S'_2$ based on the coupling coefficients.

### *3.1 Inverse Problem*

It is easy to determine that the coefficients of Eq. (17) are related to its roots $S'_1$ and $S'_2$ as follows

$$S'_1 + S'_2 = m_{013}/a_{13}, \tag{18a}$$

$$S'_1 S'_2 = m_{12}m_{23}/a_{13}. \tag{18b}$$

Equalities (18a, 18b) represent the inverse problem. We assume that the coefficients $m_{12}$, $m_{23}$ are known. Based on (18a, 18b), it is required to determine the mixed cross coupling $m_{13}(\Omega) = m_{013} - a_{13}\Omega$. The solution of the inverse problem is

$$a_{13} = m_{12}m_{23}/S'_1 S'_2, \tag{19a}$$

$$m_{013} = m_{12}m_{23}\left(S'_1 + S'_2\right)/S'_1 S'_2\,. \tag{19b}$$

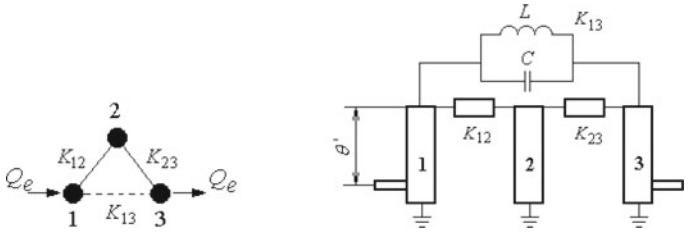



**Fig. 1** Trisection BPF with mixed cross coupling: **a** Coupling structure; **b** Schematic

Consider some particular cases of solution (19a, 19b). Let the transmission zeros be located equidistantly on the $\sigma$ axis: $S_1 = \sigma_1$ ($S_1' = \sigma_1/j$); $S_2 = -\sigma_1$($S_2' = -\sigma_1/j$). It follows from (19b) that $m_{013} = 0$ ($K_{m13} = |K_{e13}|$). The equality (19a) is satisfied if the main coupling coefficients have the same sign ($m_{23} = m_{12}$). In this case it takes the form

$$a_{13} = K_{m13} = |K_{e13}| = m_{12}^2/\sigma_1^2. \tag{20}$$

The equidistant arrangement of transmission zeros on the $\sigma$ axis is ensured in the symmetric ($K_{12} = K_{23}$) trisection BPF with mixed cross coupling $K_{13} = 0$.

If transmission zeros are located equidistantly on the $j\Omega$ axis: $S_1 = j\Omega_1$ ($S_1' = \Omega_1$), $S_2 = -j\Omega_1$ ($S_2' = -\Omega_1$), then we obtain $m_{013} = 0$ ($K_{m13} = |K_{e13}|$) from (19b). The equality (19a) is satisfied if the main coupling coefficients have different signs ($m_{23} = -m_{12}$). In this case we obtain

$$a_{13} = K_{m13} = |K_{e13}| = m_{12}^2/\Omega_1^2. \tag{21}$$

The equidistant arrangement of transmission zeros on the $j\Omega$ axis is provided in the asymmetric ($K_{12} = -K_{23}$) trisection BPF with $K_{13} = 0$. Expression (21) was obtained in [19] empirically.

If $S_1 = S_2 = j\Omega_1$ ($S_1' = S_2' = \Omega_1$), then the equality (19a) is satisfied if the main coupling coefficients have the same sign ($m_{23} = m_{12}$). In this case, equalities (19a, 19b) take the form

$$a_{13} = m_{12}^2/\Omega_1^2, \tag{22a}$$

$$m_{013} = 2m_{12}^2/\Omega_1. \tag{22b}$$

The transmission zero of the second degree on the $j\Omega$ axis is provided in the symmetric ($K_{12} = K_{23}$) trisection BPF. The expression (22a, 22b) shows that the convergence of transmission zero to center frequency $f_0$ leads to an increase in the values of $m_{013}$ and $a_{13}$.

### *3.2 Various Arrangements of Transmission Zeros*

Using the inverse problem solution (19a, 19b), we sequentially establish a complete set of all possible placements of transmission zeros of the filter under consideration. To do this, consider the trisection BPF scheme shown in Fig. 1b.

The filter used λ/4 resonators with characteristic impedances $Z_0 = 10\ \Omega$, susceptance slope parameter $b = \pi/4Z_0$ [45], and $f_0 = 1$ GHz. We use segments of quarter-wave transmission lines with characteristic impedance $Z_{i,i+1}$ as $J$-inverters. In this case, the coupling coefficient between resonators is

$$K_{i,\,i+1} = 4Z_0 \big/ \pi\ Z_{i,\,i+1}. \tag{23}$$

To implement a mixed coupling, an $LC$ circuit is used, the elements of which are connected with the components $K_m$, $K_e$ by the relationships [33]:

$$K_m \approx 2Z_0 \big/ \pi^2 f_0 L; \tag{24a}$$

$$K_e \approx -8Z_0 f_0 C \tag{24b}$$

Suppose a Chebyshev prototype with a 0.1 dB ripple, $FBW = 0.05$, and $f_0 = 1$ GHz is chosen. The prototype parameters are $g_0 = 1$, $g_1 = g_3 = 1.0315$, $g_2 = 1.1474$. Using formulas [45] we determine the main coupling coefficients $K_{12} = K_{23} = 0.0420$ ($m_{12} = 0.84$) and the external quality factor $Q_e = 24.55$. From (23) we determine that the above coupling coefficients are provided at $Z_{12} = Z_{23} = 303.15\ \Omega$. The filter loads $R_L = 50\ \Omega$ are conductively connected to the end (λ/4) resonators. The required value of the external quality factor $Q_e$ is provided by choosing the coordinate $\theta'$, measured from the open end of the λ/4 resonator $Q_e = R_L b = R_L \pi/4Z_0 \cos^2\theta'$. From this equation we determine the value $Q_e = 24.55$, which is provided at the connection coordinate $\theta' = 66.42°$.

The various options for placing the transmission zeros of the filter under consideration and the corresponding frequency responses are presented in Table 1. The first seven positions are implemented in a symmetrical trisection BPF ($K_{12} = K_{23}$).

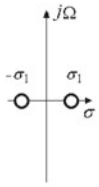
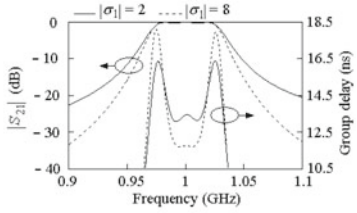
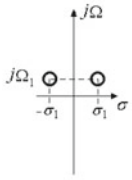
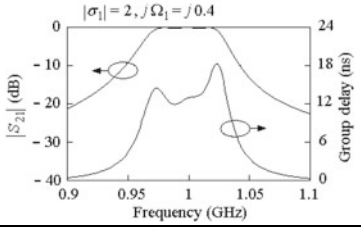
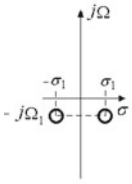
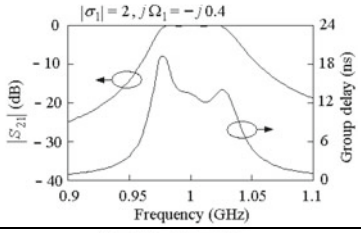
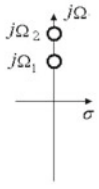
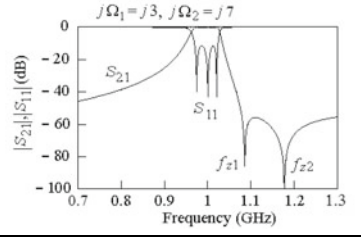
**Case 1.** Trisection BPF with equidistant transmission zeros placed on the $\sigma$ axis (No. 1 in Table 1). In this case $K_{13} = 0$. Assuming $|\sigma_{1,2}| = 8$, we obtain $a_{13} = K_{m13} = |K_{e13}| = 0.01102$ from (20). According to (24a), the value $K_{m13} = 0.01102$ is realized by inductance $L = 183.8$ nH. The value $K_{e13} = -0.01102$ is realized by the capacitance $C = 0.1378$ pF (24b). In position No. 1 the dotted lines depict the frequency response of the filter, simulated using by linear simulator of the Microwave Office (AWR). Group delay of the filter is $\tau = 11.8 \pm 0.1$ ns. Let us set the value $|\sigma_{1,2}| = 2$. As a calculation result we obtain: $a_{13} = K_{m13} = |K_{e13}| = 0.1764$; $L = 11.488$ nH; $C = 2.205$ pF. Solid lines in the position No.1 depict BPF frequency responses for this case. We get the group delay $\tau = 13.3 \pm 0.15$ ns. The convergence of transmission zeros by four times led to an increase in 16 times the components $K_m = |K_e|$ of the mixed coupling. The condition $K_{13} = 0$ for flat group delay, was presented in [33].

**Case 2.** Complex transmission zeros located in the upper half-plane $S$ (No. 2 in Table 1). We use the transmission zeros $S_1 = 2 + j0.4$ ($S_1' = 2/j + 0.4$) and $S_2 = -2 + j0.4$ ($S_2' = -2/j + 0.4$). Substituting the values into (19a, 19b) we get the mixed

cross coupling in linear presentation form: $a_{13} = 0.1696$; $m_{013} = 0.1357$. Then we find the value of the cross coupling coefficient (13a) $K_{13} = m_{013}\ FBW = 0.006785$ and using (1) we obtain the components $K_{m13} = 0.173$ ($L = 11.713$ nH) and $K_{e13} = -0.1662$ ($C = 2.0775$ pF). The parameters of the filter for this case are fully defined and its simulated frequency responses are shown in position No. 2 of Table 1.

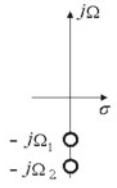
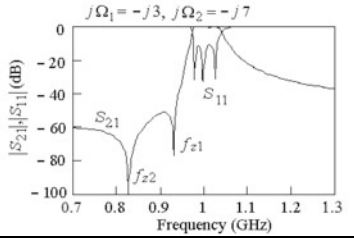
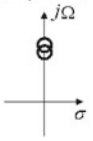
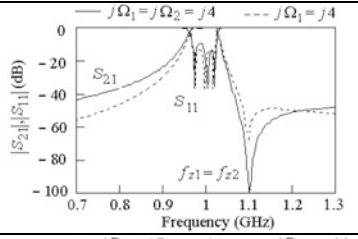
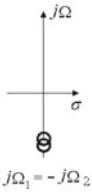
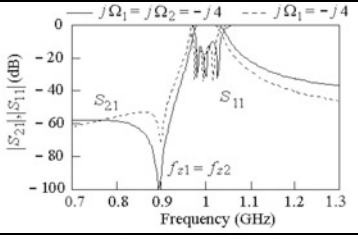
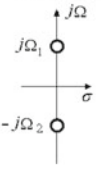
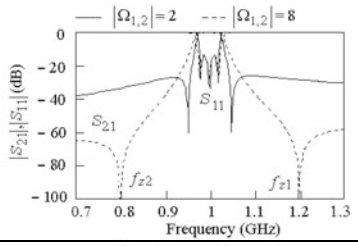
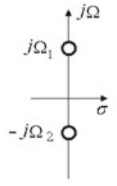
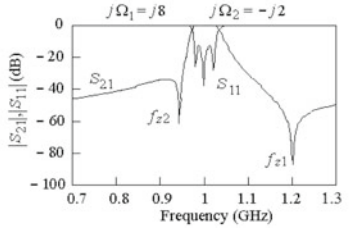
**Case 3.** Complex transmission zeros located in the lower half-plane $S$ (No. 3 in Table 1). We use the transmission zeros: $S_1 = 2 - j0.4$ and $S_2 = -2 - j0.4$. Performing calculations in the sequence set above, we get: $a_{13} = 0.1696$, $m_{013} = -0.1357$, $K_{13} = -0.006785$, $K_{m13} = 0.1662$ ($L = 12.192$ nH) and $K_{e13} = -0.1730$ ($C = 2.162$

**Table 1** Transmission zeros and frequency responses of trisection BPF with mixed cross-coupling

| Transmission zeros | Frequency responses | $K_{13}$ | $K_{12}$, $K_{23}$ |
|---|---|---|---|
| No. 1  |  | $K_{13} = 0$ | $K_{12} = K_{23}$ |
| No. 2  |  | $K_{13} > 0$ | |
| No. 3  |  | $K_{13} < 0$ | |
| No. 4  |  | $K_{13} > 0$ | |

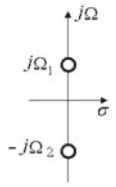
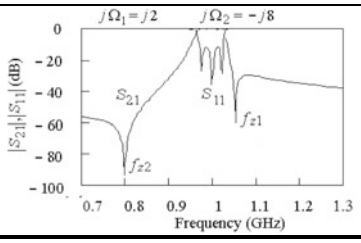
(continued)

**Table 1** (continued)

| | | | |
|---|---|---|---|
| No. 5   |   | $K_{13} < 0$ | |
| No. 6 $j\Omega_1 = j\Omega_2$   |   | $K_{13} > 0$ | $K_{12} = K_{23}$ |
| No. 7   |   | $K_{13} < 0$ | |
| No. 8 $j\Omega_2 = -j\Omega_1$   |   | $K_{13} = 0$ | $K_{12} = -K_{23}$ |
| No. 9   |   | $K_{13} > 0$ | |

(continued)

**Table 1** (continued)

| | | | |
|---|---|---|---|
| No. 10   |   | $K_{13} < 0$ | $K_{12} = -K_{23}$ |

pF). The simulated frequency responses of the filter are shown in position No. 3 of Table 1. The filters No. 2, 3 has no flat group delay if its two transmission zeros are moved from the $\sigma$ axis to the upper or lower $S$ half-plane. In [13] it is stated that the linearization effect of the phase and flat group delay will be preserved when two transmission zeros placed equidistant on the σ axis are shifted to the lower part of the complex plane $S$. The above results show that the statement [28] is not valid.

**Case 4.** Two right-sided transmission zeros at the $j\Omega$ axis (No. 4 in Table 1). Assuming $S_1 = j3$ ($f_{z1} = 1.075$ GHz), $S_2 = j7$ ($f_{z2} = 1.175$ GHz) or $S_1' = 3$ and $S_2' = 7$. Solution (19a, 19b) of the inverse problem gives: $a_{13} = 0.0336$, $m_{013} = 0.336$, $K_{13} = 0.0168$, $K_{m13} = 0.0430$ ($L = 47.126$ nH) and $K_{e13} = -0.0262$ ($C = 0.3275$ pF). The simulated frequency responses of the filter are shown in position No. 4 of Table 1.

**Case 5.** Two left-sided transmission zeros at the $j\Omega$ axis (No. 5 in Table 1). Assuming $S_1 = -j3$ ($f_{z1} = 0.925$ GHz), $S_2 = -j7$ ($f_{z2} = 0.825$ GHz). From solution (19a, 19b) we get: $a_{13} = 0.0336$; $m_{013} = -0.336$, $K_{13} = -0.0168$; $K_{m13} = 0.0262$ ($L = 77.344$ nH) and $K_{e13} = -0.043$ ($C = 0.5375$ pF). Simulated frequency responses of the filter are shown in position No. 5 of Table 1.

**Case 6.** Right-sided second-order transmission zero at the $j\Omega$ axis (No. 6 in Table 1). Assuming $S_1 = S_2 = j4$ ($f_z = 1.100$ GHz), or $S_1' = S_2' = 4$. Substituting the values into (22a, 22b) we get: $a_{13} = 0.0441$, $m_{013} = 0.3528$. Expressions (13a, 13b) give the values: $K_{13} = 0.01764$, $K_{m13} = 0.0538$ ($L = 37.665$ nH), $K_{e13} = -0.0362$ ($C = 0.4525$ pF). The simulated frequency responses for this case are shown in position No. 6 of Table 1 with the solid line. In the same figure, the dashed line shows the frequency responses of a trisection BPF with a simple transmission zero.

**Case 7.** Left-sided second-order transmission zero at the $j\Omega$ axis (No. 7 in Table 1). Assuming $S_1 = S_2 = -j4$ ($f_z = 0.9$ GHz). Expressions (13a, 13b) give the values: $a_{13} = 0.0441$, $m_{013} = -0.3528$, $K_{13} = -0.01764$, $K_{m13} = 0.0362$ ($L = 55.978$ nH), $K_{e13} = -0.0538$ ($C = 0.6725$ pF). The simulated frequency responses for this case are shown in position No. 7 of Table 1 with the solid line. The dashed line shows the frequency responses of BPF with a simple transmission zero. The second-order transmission zero provides a deeper attenuation pole.

From (17), (22a, 22b) it follows that the second-order transmission zero is provided when the condition holds:

$$K_{13}^2/4a_{13} = K_{12}^2. \tag{25}$$

The second-order transmission zero is located at the frequency

$$f_z = f_0\left[1 + \left(\sqrt{K_m/|K_e|} - \sqrt{|K_e|/K_m}\Big/4\right)\right]. \tag{26}$$

Case 7 corresponds to the following values: $K_{12} = 0.042$, $K_{13} = -0.01764$, $a_{13} = 0.0441$. Substitution them into (25) shows that this equality holds. Substituting other values $f_0 = 1$ GHz, $K_{m13} = 0.0362$, $K_{e13} = -0.0538$ in (26), we determine the transmission zero $f_z = 0.9$ GHz, which coincides with the displayed in position No. 7 of Table 1.

The next three positions from No. 8 to No. 10 are implemented in an asymmetrical trisection BPF ($K_{12} = -K_{23}$).

**Case 8.** Equidistant transmission zeros at the $j\Omega$ axis (No. 8 in Table 1). If $|\Omega_{1,2}| = 8$, then using (21), (24a, 24b) we obtain: $K_{m13} = 0.01102$ ($L = 183.8$ nH), $K_{e13} = -0.01102$ ($C = 0.1378$ pF), $K_{13} = 0$. In position No. 8 the dotted lines depict the frequency response of the filter. Let us set the value $|\Omega_{1,2}| = 2$. As a calculation result we obtain: $K_{m13} = 0.1764$ ($L = 11.488$ nH), $K_{e13} = -0.1764$ ($C = 2.205$ pF), $K_{13} = 0$. Solid lines in the same figure depict BPF frequency responses for this case. The convergence of transmission zeros by four times (from $|\Omega_{1,2}| = 8$ to $|\Omega_{1,2}| = 2$) led to an increase in 16 times the components $K_m = |K_e|$.

**Case 9.** Two-sided transmission zeros at the $j\Omega$ axis, shifted upward in the frequency (No. 9 in Table 1).

Assuming $S_1 = j8$ ($f_{z1} = 1.2$ GHz), $S_2 = -j2$ ($f_{z2} = 0.95$ GHz) or $S_1' = 8$ and $S_2' = -2$. Formulas (19a, 19b), (13a, 13b) give: $a_{13} = 0.0441$, $m_{013} = 0.2646$, $K_{13} = 0.01323$, $K_{m13} = 0.0512$ ($L = 39.58$ nH) and $K_{e13} = -0.038$ ($C = 0.475$ pF). Frequency responses of the BPF with predicted transmission zeros are shown in position No. 9 of Table 1.

**Case 10.** Two-sided transmission zeros at the $j\Omega$ axis, shifted downward in the frequency (No. 10 in Table 1).

If $S_1 = j2$ ($f_{z1} = 1.05$ GHz), $S_2 = -j8$ ($f_{z2} = 0.8$ GHz), then using (19a, 19b), (13a, 13b) we obtain: $a_{13} = 0.0441$, $m_{013} = -0.2646$, $K_{13} = -0.01323$, $K_{m13} = 0.038$ ($L = 53.33$ nH) and $K_{e13} = -0.0512$ ($C = 0.64$ pF). Frequency responses of the BPF are shown in position No. 10 of Table 1.

The trisection BPF under consideration has 10 different placement options for two transmission zeros (Table 1). Other placements are physically unrealizable, since they do not lead to real coefficients (18a, 18b) of polynomial (17) of the considered trisection BPF. The existence of a second-order transmission zero (No. 6, 7) in the

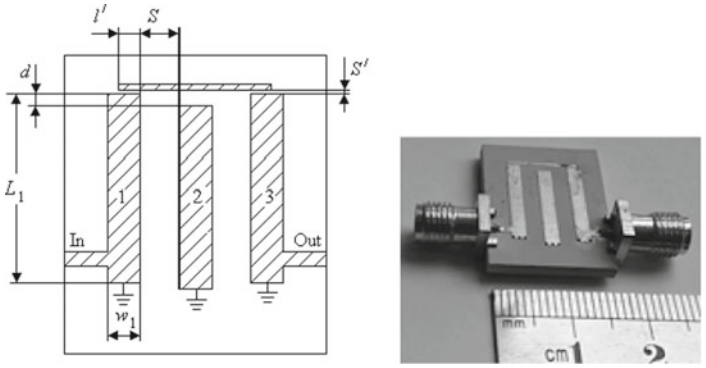
trisection BPF with mixed $K_{13}$ coupling was established for the first time, which provides a deeper attenuation pole at insertion loss frequency response.

Ten options for two transmission zeros placing in trisection BPFs with mixed cross coupling can be confirmed experimentally using two symmetrical microstrip filters.

## 3.3 Microstrip Quasi-Inline Trisection BPFs with λ/4 Resonators

Figures 2a, 2b show the topology and photograph of the microstrip quasi-inline trisection BPF with λ/4 resonators, which occupies an area of 21.6 mm × 16.8 mm. The filter uses the substrate TMM-10i (Rogers), $\varepsilon_r = 9.8$, $\tan\delta = 0.002$, $h = 1.905$ mm. Short circuit at the ends of the resonators is provided by metalized holes with the diameter 0.4 mm. Center resonator *2* is shifted downward relative to the resonators *1* and *3* for the distance $d = 0.9$ mm, due to this the inline filter becomes the quasi-inline filter. Resonant frequency of the resonator is 2 GHz. The cross coupling $K_{13}$ between resonators *1* and *3* is a mixed one. Its magnetic component $K_{m13}$ is a parasitic coupling between these resonators. These filter parameters correspond to the coupling coefficients $K_{12} = K_{23} = 0.056$, $K_{m13} = 0.0139$. The procedure for determining the coupling coefficients is well known [34]. For this purpose, the EM simulator of computer program Microwave Office (AWR) is used.

The coordinate of connecting the I/O line gives the external $Q$-factor of the end resonators $Q_e = 17.06$. The strip connecting resonators *1* and *3* through the capacitive gaps insert the electrical component $K_{e13}$ into the mixed cross coupling. An increase in $|K_{e13}|$ occurs with an increase in the length $l'$ and with a decrease in the gap $S'$. If




a) b)

**Fig. 2** Microstrip quasi-inline trisection BPF with λ/4 resonators: **a** Topology: $L_1 = 13.6$ mm, $w_1 = 2$ mm, $S = 2.4$ mm, $d = 0.9$ mm, $S' = 0.2$ mm, $l' = 1.8$ mm; **b** Photograph of fabricated filter

$S' = 0.2$ mm, $l' = 1.8$ mm, than $|K_{e13}| = K_{m13} = 0.0139$, $K_{13} = 0$. The transmission zeros of this filter are located equidistantly on the $\sigma$-axis and the filter has the flat group delay.

The measured and simulated frequency responses for this filter with two different geometric parameters are shown in Table 2, position No. 1. The insets show the filter topologies. The measured data are as follows: $f_0 = 2$ GHz, BW = 151 MHz, $IL_0 = 1.0$ dB; $RL > 14.2$ dB; group delay $\tau = 4.3 \pm 0.1$ ns within the frequency range of 74 MHz. If we use geometric parameters $S = 1$ mm and $d = 2.7$ mm then the value of $K_{12}$ preserved, while the parasitic magnetic cross coupling increased $K_{m13} = 0.0271$. To compensate this value of $K_{m13}$, the coupling strip with the parameters $S' = 0.05$ mm, $l' = 1.8$ mm is used. The group delay increases from 4.3 ns to 4.9 ns.

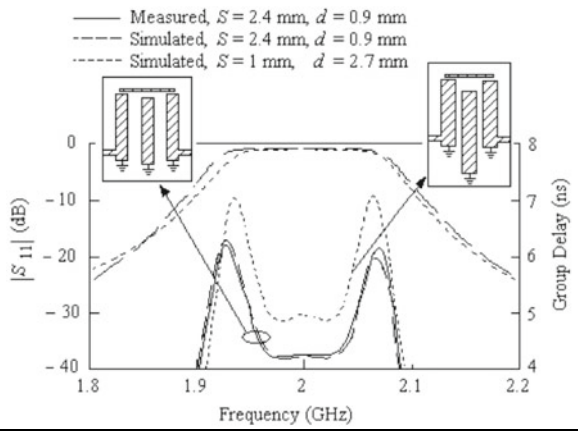
The considered trisection BPF (Fig. 2) is symmetrical ($K_{12} = K_{23}$) and it has 7 different options for the transmission zeros placement and their corresponding frequency responses (No. 1–No. 7, Table 2). The position numbers in Table 2 are the same as those in Table 1. Frequency responses No. 1 correspond to mixed cross-coupling $K_{13} = 0$. Frequency responses No. 2, 4, 6 occur at $K_{13} > 0$, and frequency responses No. 3, 5, 7 are provided at $K_{13} < 0$. To implement positive values of $K_{13}$, the gap $S' = 0.2$ mm was used, and negative values of $K_{13}$ were provided at $S' = 0.025$ mm. In all cases, the filter parameters $S = 2.4$ mm, $d = 0.9$ mm and $K_{m13} = 0.0139$ unchanged. Only the electrical component $K_{e13}$ was changed by changing the value of $l'$.

The frequency responses of positions No. 2 and No. 3 correspond to the cases of displacement of equidistant transmission zeros on the $\sigma$-axis to the upper (No. 2) or lower (No. 3) half-plane $S$. In both cases, the flat group delay is not provided.

The frequency responses of positions No. 4 and No. 5 are characterized by two one-sided transmission zeros at the real frequencies. In the first case (No. 4) at $K_{13} >$

**Table 2** Frequency responses of microstrip quasi-inline trisection BPF with mixed cross-coupling

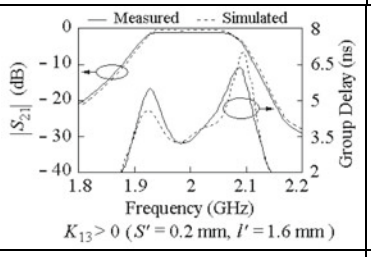
No. 1




(continued)

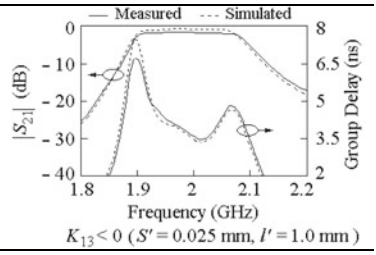
**Table 2** (continued)

No. 2



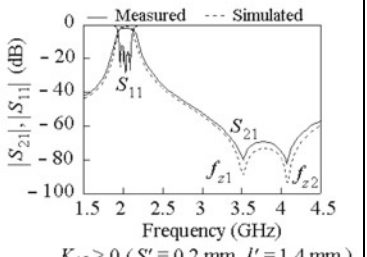

$K_{13} > 0$ ($S' = 0.2$ mm, $l' = 1.6$ mm)

No. 3



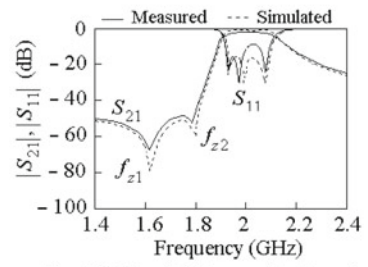

$K_{13} < 0$ ($S' = 0.025$ mm, $l' = 1.0$ mm)

No. 4



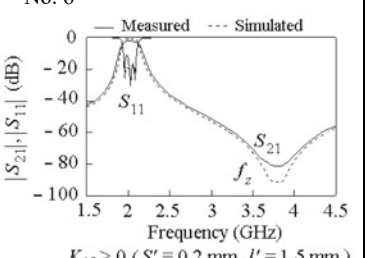

$K_{13} > 0$ ($S' = 0.2$ mm, $l' = 1.4$ mm)

No. 5



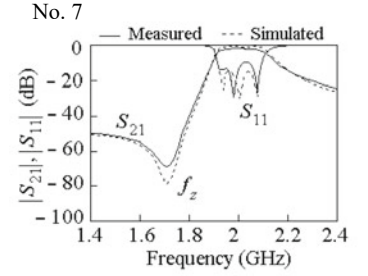

$K_{13} < 0$ ($S' = 0.025$ mm, $l' = 2$ mm)

No. 6



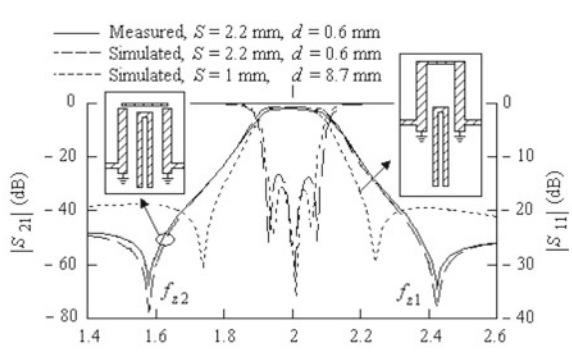

$K_{13} > 0$ ($S' = 0.2$ mm, $l' = 1.5$ mm)

No. 7



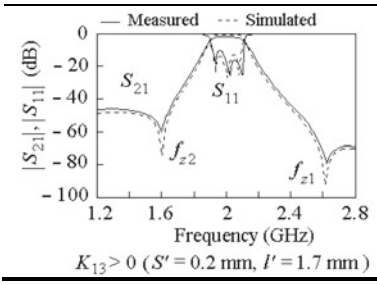
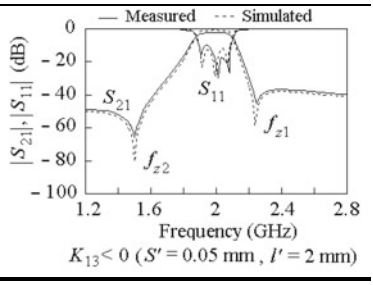

$K_{13} < 0$ ($S' = 0.025$ mm, $l' = 1.8$ mm)

No. 8




(continued)

**Table 2** (continued)

| No. 9 | No. 10 |
|---|---|
|  |  |
| $K_{13} > 0$ ($S' = 0.2$ mm, $l' = 1.7$ mm) | $K_{13} < 0$ ($S' = 0.05$ mm, $l' = 2$ mm) |



0, the transmission zeros $f_{z1} = 3.51$ GHz and $f_{z2} = 4.08$ GHz are located to the right of $f_0$. In the second case (No. 5) at $K_{13} < 0$, the transmission zeros $f_{z1} = 1.798$ GHz and $f_{z2} = 1.618$ GHz are located to the left of $f_0$.

The frequency responses with one second-order transmission zero occupy positions No. 6 and No. 7. At $K_{13} > 0$, the transmission zero $f_z = 3.8$ GHz is right-handed (No. 6). At $K_{13} < 0$, the transmission zero $f_z = 1.708$ GHz is located to the left of $f_0$ (No. 7).

## *3.4 Microstrip Quasi-Inline Trisection BPFs with λ/4 and λ/2 Resonators*

The locations of the two transmission zeros corresponding to the three cases 8–10 (Table 1) require the use of asymmetric ($K_{12} = -K_{23}$) trisection BPF, which is not always convenient to use. However, these BPFs can be made symmetrical [34] if the central reflection-type quarter-wave resonator (Fig. 3a) is replaced by a though-type resonator (λ/2), as shown in Fig. 3b by a double circle. In addition to the resonance phenomenon, the through-type resonator implements the transformation of impedances and changes the sign of the coupling coefficient. Replacing a reflection-type resonator with a though-type resonator significantly changes the frequency response of the BPF, so these resonators require different designations in the coupling structure.

Figures 4a, b show the topology and photograph of the microstrip quasi-inline trisection BPF with λ/4 and λ/2 resonators, which occupies an area of 21.6 mm × 17.4 mm. The parameters of quarter-wave resonators and substrate are the same as in the previous filter. Hairpin resonator *2* is a through-type resonator with the microstrip line of width 1 mm. This resonator is shifted downward relative to the resonators *1* and *3* for the distance $d = 0.6$ mm. These filter parameters correspond to the main coupling coefficients $K_{12} = K_{23} = 0.065$, and magnetic component of mixed cross coupling $K_{m13} = 0.0235$. The coordinate of connecting the I/O line gives the

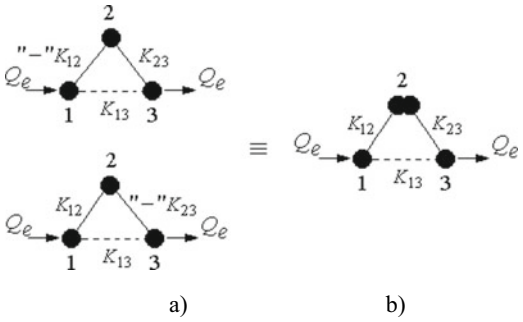



**Fig. 3** Coupling structures of trisection BPF: **a** Asymmetric; **b** Symmetric

external $Q$-factor of the end resonators $Q_e = 14.67$. The electrical component $K_{e13}$ of the mixed coupling $K_{13}$ is determined by the $S'$ and $l'$ values of the strip connecting resonators *1* and *3* through the capacitive gaps. If $S' = 0.1$ mm and $l' = 2.0$ mm, than $|K_{e13}| = K_{m13} = 0.0235$, $K_{13} = 0$ and transmission zeros are located equidistantly on the $j\Omega$ axis. It corresponds to a quasi-elliptic filter.

The measured and simulated frequency responses for this filter with two different geometric parameters are shown in Table 2, position No. 8. The insets show the filter topologies. The measured data are as follows: the center frequency $f_0 = 2$ GHz, $BW = 171$ MHz, midband insertion loss $IL_0 = 0.9$ dB; return loss in the bandpass $RL > 13$ dB. Transmission zeros $f_{z1} = 2425$ MHz and $f_{z2} = 1580$ MHz are almost equidistant to $f_0$: $f_{z1} - f_0 = 425$ MHz; $f_0 - f_{z2} = 420$ MHz. The calculated value of $|f_{z1,2} - f_0|$ by formula (21) is $|f_{z1,2} - f_0| = 424$ MHz. The selectivity of this filter 35 dB ($f_0 \pm 280$ MHz) can be expressed at the normalized frequency $j\Omega$: 35 dB

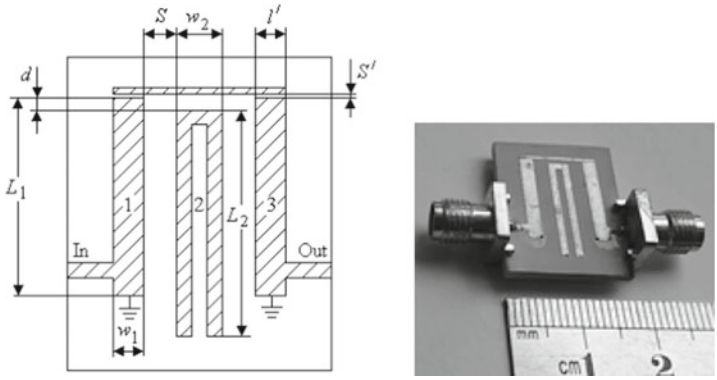



**Fig. 4** Microstrip quasi-inline trisection BPF with λ/4 and λ/2 resonators: **a** Topology: $L_1 = 13.6$ mm, $w_1 = 2$ mm, $L_2 = 15.7$ mm, $w_2 = 3$ mm, $S = 2.2$ mm, $d = 0.6$ mm, $S' = 0.1$ mm, $l' = 2$ mm; **b** Photograph of fabricated filter

($\pm j3.3$). This selectivity is higher than the third order "Ceramic Monoblock Filter CER0003A" [46]: 34 dB ($-j4.66$); 16.5 dB ($j4.6$).

If we use geometric parameters $S = 1$ mm and $d = 8.7$ mm then the value of $K_{12}$ preserved, while the parasitic magnetic cross coupling increased $K_m = 0.0385$. To compensate this value of $K_m$, the coupling strip with the parameters $S' = 0.05$ mm, $l' = 2$ mm is used. Frequency responses at position No. 8 (Table 2) show the movement of transmission zeros by changing geometric parameters of this filter.

The frequency responses of positions No. 9 and No. 10 (Table 2) are characterized by two versatile transmission zeros at the real frequencies. In the first case (No. 9) at $K_{13} > 0$, the transmission zeros $f_{z1} = 2.616$ GHz and $f_{z2} = 1.60$ GHz are shifted upward in frequency. In the second case (No. 10) at $K_{13} < 0$, the transmission zeros $f_{z1} = 2.237$ GHz and $f_{z2} = 1.50$ GHz are shifted downward in frequency.

## 4 Transmission Zeros of Quadruplet BPF with Mixed Cross-Coupling

Consider a symmetrical quadruplet BPF with $K_{12} = K_{34}$ ($m_{12} = m_{34}$), its coupling structure is shown in Fig. 5a. If this BPF uses a mixed cross-coupling $K_{14}$, then we replace $m_{14}$ with $m_{14}(S') = m_{014} - a_{14}S'$ in (6)

$$M_{41} = \left(m_{014} - a_{14}S'\right)S'^2 + m_{12}^2 m_{23} - \left(m_{014} - a_{14}S'\right)m_{23}^2. \tag{27}$$

Equating minor (27) to zero, we obtain the algebraic equation of the third degree defining the transmission zeros

$$S'^3 - \frac{m_{014}}{a_{14}}S'^2 - m_{23}^2 S' - \frac{m_{12}^2 m_{23}}{a_{14}} = 0. \tag{28}$$

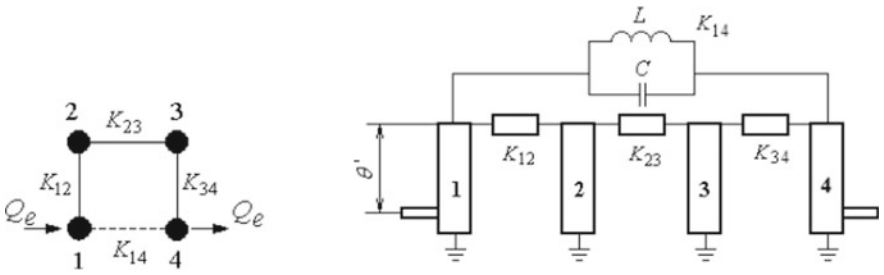



**Fig. 5** Quadruplet BPF with mixed cross coupling: **a** Coupling structure; **b** Schematic

The location of the roots of polynomial (28) depends on the relationship between the signs of the coupling coefficients $m_{014}$ and $m_{23}$.

## 4.1 Inverse Problem

The inverse problem (16) for the considered quadruplet BPF described by (28) is represented by a system of three algebraic equations regarding $m_{14}(\Omega) = m_{014} - a_{14}$ $\Omega$:

$$-S_1' S_2' S_3' = -m_{12}^2 m_{23}/a_{14}, \tag{29a}$$

$$-S_1' S_2' S_3' = -m_{12}^2 m_{23}/a_{14}, \tag{29b}$$

$$S_1' \prime S_2' + S_1' S_3' + S_2' S_3' = -m_{23}^2. \tag{29c}$$

System (29a, 29b) is solved by the equalities

$$a_{14} = m_{12}^2 m_{23}/S_1' S_2' S_3', \tag{30a}$$

$$m_{014} = m_{12}^2 m_{23}\left(S_1' + S_2' + S_3'\right)/S_1' S_2' S_3', \tag{30b}$$

$$S_3' = -\left(m_{23}^2 + S_1' S_2'\right)/\left(S_1' + S_2'\right). \tag{30c}$$

Equality (30c) does not contain the determinable quantities $m_{014}$ and $a_{14}$, it indicates that the three transmission zeros of quadruplet BPF with the mixed cross coupling are interdependent. We can specify only two of them, and the third transmission zero must be determined from the condition (30c). When obtaining equality (30c), we assume that transmission zeros $S_1'$ and $S_2'$ are independent. From solution (30a, 30b, 30c), the restrictions follow for some variants of the placement of three transmission zeros:

- None of the transmission zeros can be placed on the $S = \sigma$ ($S' = j\sigma$) axis, otherwise the value $m_{014}$ (30b) will not be real.
- Equidistant arrangement of two transmission zeros on the $S = \sigma$ ($S' = j\sigma$) axis is also not allowed. In this case, the denominator of the right side (30c) will be equal to zero, which is not allowed.
- Three transmission zeros on the $S = \sigma(S' = j\sigma)$ axis cannot be one-sided. If two transmission zeros have the same sign, then the third transmission zero should have the opposite sign (30c).

- Two of the three transmission zeros on the $S = j\Omega (S' = \Omega)$ axis cannot be equidistant. If $j\Omega_2 = -j\Omega_1$ ($\Omega_2 = -\Omega_1$), then the denominator of the right side (30c) will be equal to zero, which is not allowed.

## 4.2 *Various Arrangements of Transmission Zeros*

Using the inverse problem solution (30a, 30b, 30c), we establish the various admissible options for the transmission zeros placement of a quadruplet BPF with mixed cross coupling $K_{14}$.

Consider the quadruplet BPF scheme (Fig. 5b), which contains the same elements as the trisection BPF (Fig. 1b). Suppose a Chebyshev prototype with a 0.2 dB ripple, $N = 4$, $FBW = 0.05$, and $f_0 = 1$ GHz is chosen. The prototype parameters are $g_0 = 1$, $g_1 = 1.3028$, $g_2 = 1.2814$, $g_3 = 1.9761$, $g_4 = 0.8468$. Using formulas [45] we determine $K_{12} = K_{34} = 0.0387$ ($m_{12} = 0.774$), $|K_{23}| = 0.0314$ ($|m_{23}| = 0.628$), $Q_e = 26.056$. These coupling coefficients correspond to the characteristic impedances of the quarter-wave segments connecting between the resonators (23): $Z_{12} = Z_{34} = 329\ \Omega$; $|Z_{23}| = 409.49\ \Omega$. Expressions (28), (29a, 29b, 29c) show that the signs of the main coupling coefficients $K_{12} = K_{34}$ do not affect the transmission zeros, and we assigned them the "+" sign. The sign of $K_{23}$ has a significant effect on the transmission zeros, so we determined the modulus of this value. The load connection coordinate $\theta' = 67.15°$ provides the external quality factor of the end resonators $Q_e = 26.056$.

The various options for the transmission zeros placing of the filter under consideration and the corresponding frequency responses are presented in Table 3.

**Case 1.** Transmission zeros are placed at $S = j\Omega$ ($S' = \Omega$) axis. Two of them are located in the lower half-plane of $S$ (No. 1 in Table 3). Let $S'_1 = 2$ ($f_{z1} = 1.05$ GHz), $S'_2 = -8$ ($f_{z2} = 0.8$ GHz). From (30c) we get $S'_3 = -2.6$ ($f_{z3} = 0.935$ GHz). Using the formulas (30a), (30b) and (13a, 13b) we obtain: $a_{14} = 0.009044$, $m_{014} = -0.0778$, $K_{14} = -0.0039$, $K_{m14} = 0.0073$ ($L = 277.59$ nH) and $K_{e14} = -0.0112$ ($C = 0.14$ pF). The simulated frequency responses for this case are shown in position No. 1.

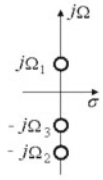
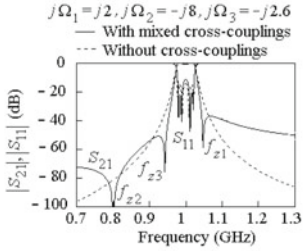
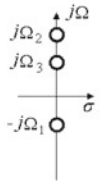
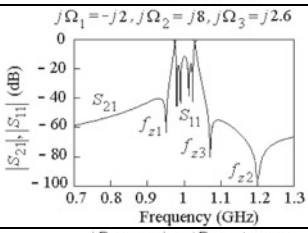
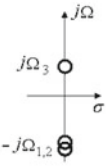
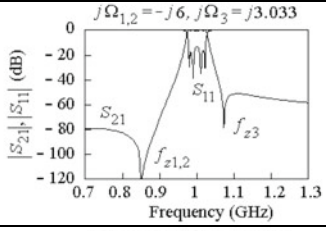
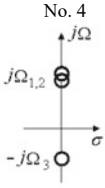
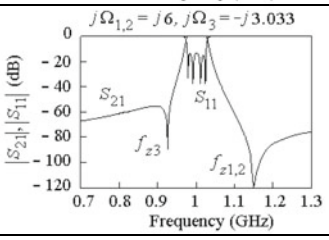
**Case 2.** Transmission zeros are located at $S = j\Omega$ ($S' = \Omega$) axis. Two of them are located in the upper half-plane of $S$ (No. 2, Table 3). Let $S'_1 = -2$ ($f_{z1} = 0.95$ GHz), $S'_2 = 8$ ($f_{z2} = 1.2$ GHz). From (30c) we find $S'_3 = 2.6$ ($f_{z3} = 1.065$ GHz). Expressions (30a), (30b) and (30c) give us: $a_{14} = 0.009044$, $m_{014} = 0.0778$, $K_{14} = 0.0039$, $K_{m14} = 0.0112$ ($L = 180.9$ nH) and $K_{e14} = -0.0073$ ($C = 0.091$ pF). The simulated frequency responses for this case are shown in position No. 2.

**Case 3.** Transmission zeros are presented at $S = j\Omega$ ($S' = \Omega$) axis. One second-order transmission zero is located in the lower half-plane of $S$ (No. 3, Table 3). Let $S'_1 = S'_2 = -6$ ($f_z = 0.85$ GHz). The solution to the inverse problem (30a, 30b, 30c) is written in the form: $S'_3 = 3.033$ ($f_{z3} = 1.0758$ GHz), $a_{14} = 0.003445$, $m_{014} = -0.0309$, $K_{14}$

$= -0.001545$, $K_{m14} = 0.00276$ ($L = 734.2$ nH) and $K_{e14} = -0.0043$ ($C = 0.05375$ pF). The simulated frequency responses for this case are shown in position No. 3.

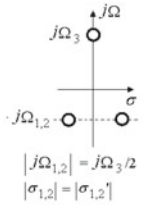
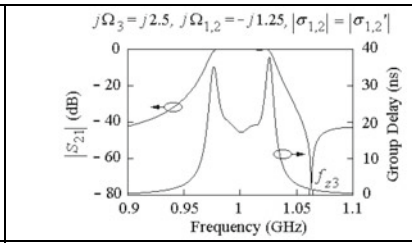
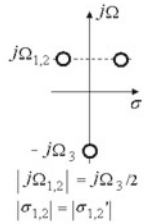
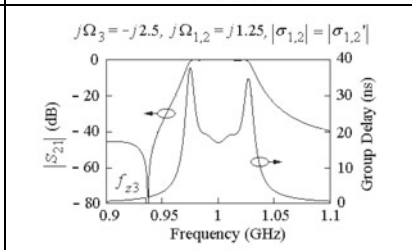
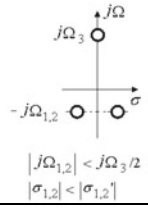
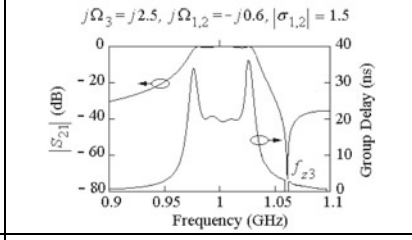
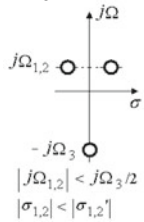
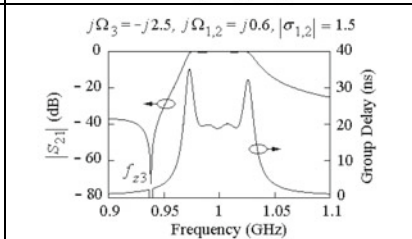
**Case 4.** Transmission zeros are placed at $S = j\Omega$ ($S' = \Omega$) axis. One second-order transmission zero is located in the upper half-plane of $S$ (No. 4, Table 3) Let $S'_1 = S'_2 = 6$ ($f_z = 1.15$ GHz). Formulas (30a, 30b, 30c), (13a, 13b) lead to the result: $S'_3 = -3.033$ ($f_{z3} = 0.9242$ GHz), $a_{14} = 0.003445$, $m_{014} = 0.0309$, $K_{14} = 0.001545$, $K_{m14} = 0.0043$ ($L = 471.26$ nH) and $K_{e14} = -0.00276$($C = 0.0345$ pF). The simulated frequency responses for this case are shown in position No. 4.

**Table 3** Transmission zeros and frequency responses of quadruplet BPF with mixed cross-coupling

| Transmission zeros | Frequency responses | $K_{14}$ | $K_{12}$ |
|---|---|---|---|
| No. 1   |   | $K_{14} < 0$ | $K_{12} > 0$ |
| No. 2   |   | $K_{14} > 0$ | $K_{12} < 0$ |
| No. 3   |   | $K_{14} < 0$ | $K_{12} > 0$ |
| No. 4   |   | $K_{14} > 0$ | $K_{12} < 0$ |

(continued)

**Table 3** (continued)

| | | | |
|---|---|---|---|
| No. 5<br>  |  | $K_{14} = 0$ | $K_{12} > 0$ |
| No. 6<br>  |  | $K_{14} = 0$ | $K_{12} < 0$ |
| No. 7<br>  |  | $K_{14} > 0$ | $K_{12} > 0$ |
| No. 8<br>  |  | $K_{14} < 0$ | $K_{12} < 0$ |

**Case 5.** Transmission zeros correspond to the condition $K_{14} = 0$ and $S_3 = j\Omega_3$ (No. 5, Table 3).

If $K_{14} = 0$, then the term $m_{014}S'^2/a_{14}$ in Eq. (28) disappears and it becomes the so-called reduced equation [22]. This equation has one real root $S'_3 = \Omega_3$ and two complex conjugates $S'_{1,2} = \Omega_{1,2} \pm j|\sigma_{1,2}|$. The inverse problem solution (30a, 30b, 30c) shows that one real root $S'_3 = \Omega_3$ defines two complex conjugate roots:

$$\Omega_{1,2} = \Omega_3/2, \tag{31}$$

$$\sigma^2_{1,2} = 3\Omega^2_3/4 - m^2_{23}. \tag{32}$$

Let $\Omega_3 = 2.5$ ($f_{z3} = 1.0625$ GHz). Then from (31), (32) we obtain $\Omega_{1,2} = -1.25$, and $|\sigma_{1,2}| = 2.329$. Since $S_1' S_2' S_3' = 17.467 > 0$, the sign of the $m_{23}$ in (30a) should be positive. Substituting the initial values into (30a), we obtain $a_{14} = K_{m14} = |K_{e14}| = 0.0215$, $L = 98.90$ nH, $C = 0.2688$ pF. The simulated frequency responses for this case are shown in position No. 5.

**Case 6.** Transmission zeros corresponding to the condition $K_{14} = 0$ and $S_3 = -j\Omega_3$ (No. 6, Table 3). Let $\Omega_3 = -2.5$ ($f_{z3} = 0.9375$ GHz). Using (31), (32) and (30a) we get: $\Omega_{1,2} = 1.25$, $m_{23} < 0$, $a_{14} = K_{m14} = |K_{e14}| = 0.0215$, $L = 98.90$ nH, $C = 0.2688$ pF. The simulated frequency responses for this case are shown in position No. 6. Note that, unlike trisection BPF, the condition $K_{14} = 0$ does not lead to a flat group delay.

**Case 7.** Transmission zeros correspond to flat group delay and $S_3 = j\Omega_3$ (No. 7, Table 3). For this, it is necessary to decrease the values of the real (31) and imaginary (32) parts of the complex transmission zeros $S_{1,2}$. Let, as in No. 5, $\Omega_3 = 2.5$ ($f_{z3} = 1.0625$ GHz), and we reduce the value $|\sigma_{1,2}| = 2.329$ to the value $|\sigma_{1,2}| = 1.5$. The value $\Omega_{1,2}$ is determined from the condition (30c): $\Omega_{1,2} = -0.6$. Note that decreasing $|\sigma_{1,2}|$ entailed a decrease in $|\Omega_{1,2}|$. The solution to the inverse problem (30a, 30b, 30c) and (13a, 13b) gives: $m_{23} > 0$, $a_{14} = 0.0809$, $m_{014} = 0.1052$, $K_{14} = 0.00526$, $K_{m14} = 0.08357$ ($L = 24.25$ nH), $K_{e14} = -0.0783$ ($C = 0.979$ pF). The simulated frequency responses for this case are shown in position No. 7.

**Case 8.** Transmission zeros correspond to flat group delay and $S_3 = -j\Omega_3$ (No. 8, Table 3). Let $\Omega_3 = -2.5$ ($f_{z3} = 0.9375$ GHz) and $|\sigma_{1,2}| = 1.5$. Carrying out the calculations in the above sequence, we obtain: $m_{23} < 0$, $a_{14} = 0.0809$, $m_{014} = -0.1052$, $K_{14} = -0.00526$, $K_{m14} = 0.0783$ ($L = 25.88$ nH), $K_{e14} = -0.08357$ (1.045 pF). The simulated frequency responses for this case are shown in position No. 8.

To obtain a flat group delay, the transmission zeros should be located at the corners of an isosceles triangle, the vertex of which lies on the $j\Omega$ axis, and the sides intersect the $\sigma$ axis. In this case, the components of the mixed cross-coupling $K_{m14}$ and $|K_{e14}|$ can be large enough for a small value of $|K_{14}|$.

### *4.3 Stripline Quadruplet BPF*

Using stripline quadruplet BPFs, we implement some positions of Table 3. Figure 6a shows the topology of stripline quadruplet BPF, which coupling structure and schematic are presented in Fig. 5. The BPF contains two SIRs (No. 1, 4) and two quarter-wave resonators (No. 2, 3) which are connected to SIRs through capacitive gaps. The main couplings $K_{12} = K_{34}$ between the resonators are negative, and the $K_{23}$ coupling is positive. The mixed cross-coupling $K_{14}$ is tuned by the parameters of SIRs [47]. It can be either positive or negative. These two cases correspond to two different states No. 7 and No. 1 from the Table 3. The substrates with parameters $h = 1$ mm, $\varepsilon_r = 36$, $\tan\delta = 0.0003$ were used in the BPF. Filter's size is 17.8 mm

$\times$ 10.2 mm $\times$ 2 mm. A resonant frequency of the resonators is $f_0 = 1.425$ GHz. The resonators are short-circuited on a metallized end surface though the metal strip with wide 1.2 mm. The length of the resonators is counted from this strip: $L_1 = 7.4$ mm; $L_2 = 7.6$ mm. The width of all resonators is the same $w = 2.4$ mm. The gaps in the filter are $S_{12} = 0.6$ mm, $S_{23} = 1.2$ mm, $S_{14} = 1$ mm. In the short-circuit region a rectangle 1.4 mm $\times$ 0.4 mm is extracted from the central strip of each SIR. Photograph of inside view of the BPF is shown in Fig. 6b. Conducting pads at the filter ends connect inner and outer parts of the filter (Fig. 6a, position 2). They are shown in the photograph of the filter (Fig. 6c), where the filter is shown on the reverse side. The input and output are coupled with end resonators by the tapped zigzag lines whose widths are equal of 0.4 mm.

With these parameters the quadruplet BPF has the main coupling coefficients $K_{12} = K_{34} = -0.0399$, $K_{23} = 0.0287$, and the mixed cross-coupling $K_{14} = 0.00294$ with components $K_m = 0.03529$, $K_e = -0.03235$. The external Q-factor $Q_e$ is determined by the position of the tapped line and is $Q_e = 23.1$. Since the signs of the coupling coefficients $K_{23}$ and $K_{14}$ coincide, and the value of $K_{14}$ is positive, this filter is a delay line filter with increased right-handed selectivity (No. 7, Table 3). The measured and

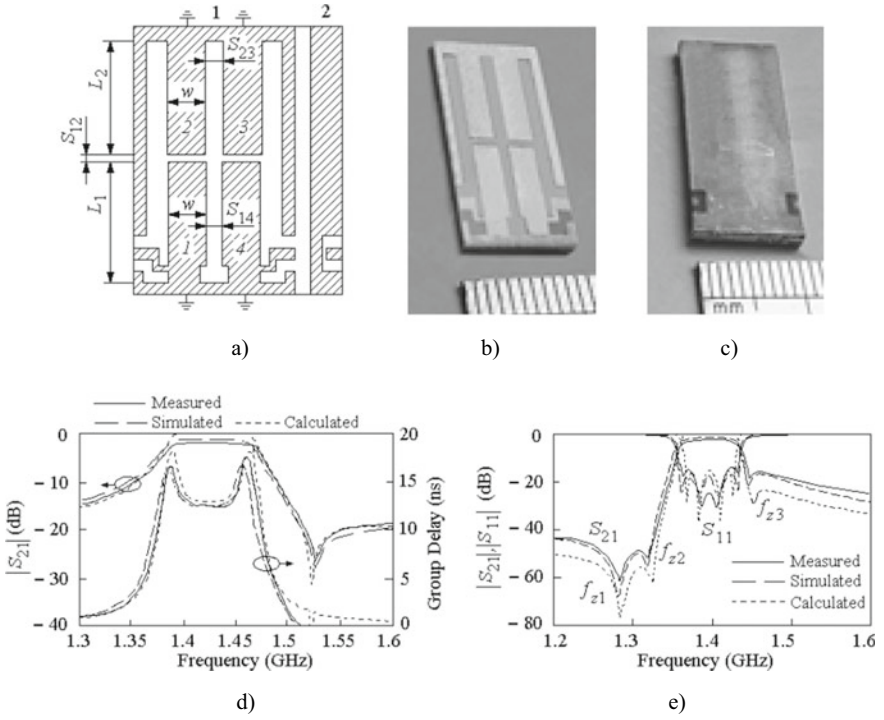



**Fig. 6** Stripline quadruplet BPF with mixed cross coupling: **a** Topology; **b** Photograph of inside view; **c** Photograph of outside view; **d** Frequency responses for case $K_{23} > 0$, $K_{14} > 0$; **e** Frequency responses for case $K_{23} > 0$, $K_{14} < 0$

simulated frequency responses for this filter are depicted in Fig. 6d. The measured parameters are as follows: $f_0 = 1425$ MHz and $BW = 72$ MHz; $IL_0 = 1.7$ dB and $RL \geq 17$ dB; the group delay $\tau = 12.7 \pm 0.2$ ns within the frequency range of 44 MHz. The transmission zero is located on the frequency $f_z = 1525$ MHz. This filter has fractional bandwidth $FBW = 0.0505$, normalized external Q-factor $q_e = 1.17$, and coupling coefficients $m_{12} = m_{34} = -0.79$, $m_{23} = 0.568$, $a_{14} = 0.0338$, $m_{014} = 0.0582$, $m_{14}(\Omega) = 0.0582 - 0.0338\Omega$. The specified values determine the normalized admittance matrix $\left[\tilde{Y}\right]$ (3) of the filter.

If we keep the basic coupling coefficients the same and change the $K_{14}$ sign to the opposite, this filter will have three transmission zeros at real frequencies (No. 1, Table 2). To change the sign of $K_{14}$, the gap $S_{14}$ between SIRs was reduced from 1 mm to 0.6 mm, and the rectangles removed from SIRs were increased from 1.4 mm $\times$ 0.4 mm to 2.0 mm $\times$ 0.5 mm. As a result, the negative value of $K_{14} = -0.015$ with the components $K_{m14} = 0.0517$ and $K_{e14} = -0.0667$ was obtained. In this case $m_{14}(\Omega) = -0.0297 - 0.0537\Omega$.

The measured and simulated frequency responses for this filter are depicted in Fig. 6e. The measured parameters are as follows: $f_0 = 1395$ MHz and $BW = 72$ MHz; $IL_0 = 1.7$ dB and $RL \geq 14$ dB. The transmission zeros are located on the frequencies $f_{z1} = 1280$ MHz, $f_{z2} = 1318$ MHz, and $f_{z3} = 1445$ MHz. The frequency $f_0$ is reduced by 30 MHz compared to the previous case (Fig. 6d), which is due to a decrease in the resonant frequency of SIRs.

## 5 Conclusion

The obtained solutions of the inverse problem for trisection BPF (19a, 19b) and quadruplet BPF (30a, 30b, 30c) with mixed cross coupling made it possible to establish restrictions on the placement of the transmission zeros of these filters, as well as to establish the new options for the location of these frequencies. It is found that the considered BPFs can have a second-order transmission zero on the $j\Omega$ axis. It is shown that to obtain a flat group delay in a quadruplet BPF, its three transmission zeros in the $S$-plane should be located at the corners of an isosceles triangle, the vertex of which lies on the $j\Omega$ axis, and the sides intersect the $\sigma$ axis. A trisection BPF with mixed cross coupling has 10 options for placing two transmission zeros, while a quadruplet BPF with simple couplings has only two such options. The solutions of the inverse problem (19a, 19b), (30a, 30b, 30c), and the establishment of new patterns in the location of the transmission zeros of these filters became possible due to the obtained expression (10) for the slope parameter $a$ in the linear representation form of the mixed coupling coefficient (8). The BPFs with mixed cross-coupling under consideration have numerous frequency responses. If two such filters are connected in cascade, then the number of realized frequency responses of the resulting BPF becomes larger. With such a connection, the influence of one filter

on another and deterioration of return loss ($S_{11}$) is possible. These distortions are restored by optimization [38, 39].

## References


1. Zolkov E, Weiss R, Cohen E (2020) Analysis and design of n-path band-pass filters with negative base band resistance. IEEE Trans Circ Syst I Reg Papers 7(67):2250–2262
2. Zhu H, Zhu X, Yang Y, Sun Y (2020) Design of miniaturized on-chip bandpass filters using inverting-coupled inductors in (Bi)-CMOS technology. IEEE Trans Circ Syst I Reg Papers 2(67):647–657
3. Zakharov A (2021) Parametric and structural-parametric synthesis of nonuniform transmission line resonators. IEEE Trans Circ Syst I Reg Papers 3(68):1055–1067
4. Zakharov A, Ilchenko M (2020) Circuit function characterizing tunability of resonators. IEEE Trans Circ Syst I Reg Papers 1(67):98–107
5. Zakharov A, Ilchenko M (2020) Unloaded quality factor of transmission line resonators with capacitors. IEEE Trans Circ Syst I Reg Papers 7(67):2204–2215
6. Makimoto M, Yamashita S (2001) Microwave resonators and filters for wireless communication. Theory, design and application. Springer
7. Fukasawa A (1982) Analysis composition of new microwave filter configuration with inhomogeneous dielectric medium. IEEE Trans Microw Theory Tech 9(MTT-30):1367–1375
8. Zakharov AV, Il'Chenko ME (2013) Thin bandpass filters containing sections of symmetric strip transmission line. J Commun Technol Electron 7(58):728–736
9. Morgan D (2010) Surface acoustic wave filters: with applications to electronic communications and signal processing. Academic Press
10. Yeung LK, Wu K-L, Wang YE (2008) Low-temperature cofired ceramic LC filters for RF applications. IEEE Microw Mag 5(9):118–128
11. Ishizaki T, Fujita M, Kagata H, Uwano T, Miyake H (1994) A very small dielectric planar filter for portable telephones. IEEE Trans Microw Theory Tech 11(MTT-42):2017–2022
12. Zakharov AV, Rozenko SA, Zakharova NA (2012) Microstrip bandpass filters on substrates with high permittivities. J Commun Technol Electron 3(57):342–351
13. Zakharov AV, Rozenko SA (2012) Duplexer designed on the basis of microstrip filters using high dielectric constant substrates. J Commun Technol Electron 6(57):649–655
14. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Hairpin resonators in varactor-tuned microstrip bandpass filter. IEEE Trans Circ Syst II Exp Briefs 10(67):1874–1878
15. Zakharov AV, Il'Chenko ME (2010) A new approach to designing varicap-tuned filters. J Commun Technol Electron 12(55):1424–1431
16. Kurzrok RM (1966) General three-resonator filters in waveguide. IEEE Trans Microw Theory Tech 1(14):46–47
17. Kurzrok RM (1966) General four-resonator filters at microwave frequencies. IEEE Trans Microw Theory Tech 6(14):295–296
18. Hershtig R, Levy R, Zaki KA (1997) Synthesis and design of cascaded trisection (CT) dielectric resonator filters. In: 27th European microwave conference digest, Jerusalem, Israel, pp 784–791
19. Li S, Huang J, Meng Q, Sun L, Zhang Q, Li F, He A, Zhang X, Li C, Li H, He Y (2007) A 12-pole narrowband highly selective high-temperature superconducting filter for the application in the third-generation wireless communications. IEEE Trans Microw Theory Tech 4(55):754–759
20. Levy R (2004) New cascaded trisections with resonant cross-couplings (CTR sections) applied to the design of optimal filters. In: IEEE MTT-S international microwave symposium digest, vol 2, pp 447–450, 6–11 June
21. Ma K, Ma J-G, Yeo KS, Do MA (2006) A compact size coupling controllable filter with separate electric and magnetic coupling paths. IEEE Trans Microw Theory Tech 3(54):1113–1119

22. Amari S, Bekheit M, Seyfert F (2008) Notes on bandpass filters whose inter-resonator coupling coefficients are linear functions of frequency. In: IEEE MTT-S international microwave symposium digest
23. Szydlowski L, Lamecki A, Mrozowski M (2012) Coupled-resonator filters with frequency-dependent couplings: coupling matrix synthesis. IEEE Microw Wireless Compon Lett 6(22):312–314
24. Szydlowski L, Jedrzejewski A, Mrozowski M (2013) A trisection filter design with negative slope of frequency-dependent cross coupling implemented in substrate integrated waveguide (SIW). IEEE Microw Wireless Compon Lett 9(23):456–458
25. Szydlowski L, Leszczynska N, Mrozowski M (2014) A linear phase filter in quadruplet topology with frequency-dependent couplings. IEEE Microw Wirel Compon Lett 1(24):32–34
26. Balewski L, Fotyga G, Lamecki A, Mrozowski M, Mul M, Sypek P, Szypulski D (2020) Step on It! Microw Mag 3(21):34–49
27. Tamiazzo S, Macchiarella G (2017) Synthesis of cross-coupled filters with frequency-dependent couplings. IEEE Trans Microw Theory Tech 3(65):775–782
28. Zhao P, Wu K (2019) Cascading fundamental building blocks with frequency-dependent couplings in microwave filters. IEEE Trans Microw Theory Tech 4(67):1432–1440
29. Shen W, Wu L-S, Sun X-W, Yin W-Y, Mao J-F (2009) Novel substrate integrated waveguide filters with mixed cross coupling (MCC). IEEE Microw Wirel Compon Lett 11(19):701–703
30. Hoft M, Shimamura T (2010) Design of symmetric trisection filters for compact low-temperature co-fired ceramic realization. IEEE Trans Microw Theory Tech 1(58):165–175
31. Wang H, Chu Q-X (2009) An inline coaxial quasi-elliptic filter with controllable mixed electric and magnetic coupling. IEEE Trans Microw Theory Tech 3(57):667–673
32. Chu Q-X, Wang H (2008) A compact open-loop filter with mixed electric and magnetic coupling. IEEE Trans Microw Theory Tech 2(56):431–439
33. Zakharov A, Ilchenko M (2017) Trisection microstrip delay line filter with mixed cross-coupling. IEEE Microw Wirel Compon Lett 12(27):1083–1085
34. Zakharov A, Rozenko S, Ilchenko M (2018) Two types of trisection bandpass filters with mixed cross-coupling. IEEE Microw Wirel Compon Lett 7(28):585–587
35. Zakharov A, Litvintsev S, Ilchenko M (2019) Trisection bandpass filters with all mixed couplings. IEEE Microw Wirel Compon Lett 9(29):592–594
36. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Trisection bandpass filter with mixed cross-coupling and different paths for signal propagation. IEEE Microw Wirel Compon Lett 1(30):12–15
37. Zakharov A, Rozenko S, Pinchuk L, Litvintsev S (2021) Microstrip quazi-elliptic bandpass filter with two pairs of anti-parallel mixed-coupled SIRs. IEEE Microw Wirel Compon Lett 5(31):433–436
38. Atia W, Zaki K, Atia A (1998) Synthesis of general topology multiple coupled resonator filters by optimization. In: IEEE MTT-S international microwave symposium dig., vol 2, pp 821–824
39. Hong J-S (2011) Microstrip filters for RF/Microwave application, 2nd edn. Wiley, Hoboken
40. Korn GA, Korn TM (1961) Mathematical handbook for scientists and engineers: definitions, theorems, and formulas for reference and review. McGraw-Hill (Inc.), New York
41. Seshu S, Balabanian N (1959) Linear network analysis. Wiley, Hoboken
42. Thomas B (2003) Cross-coupling in coaxial cavity filters-a tutorial overview. IEEE Trans Microw Theory Tech 4(51):1368–1376
43. Lu JC, Liao CK, Chang CY (2008) Microstrip parallel-coupled filters with cascade trisection and quadruplet responses. IEEE Trans Microw Theory Tech 9(56):2001–2110
44. Zakharov AV, Il'Chenko ME, Korpach VN (2014) Features of the coupling coefficients of planar stepped-impedance resonators at higher resonance frequencies and application of such resonators for suppression of spurious passbands. J Commun Technol Electron 6(59):550–556
45. Matthaei GL, Young L, Jones EMT (1980) Microwave filters, impedance-matching network, and coupling structures. Artech House, Norwood

46. CER0003A 1880 MHz PCS BPF Data Sheet, CTS Electronic Components, Bloomingdale, IL, USA (2008). https://www.ctscorp.com
47. Zakharov A, Ilchenko M (2020) Coupling coefficients between resonators in stripline combline and pseudocombline bandpass filters. IEEE Trans Microw Theory Tech 7(68):2204–2215

# Stripline Combline and Pseudocombline Bandpass Filters




**Alexander Zakharov and Michael Ilchenko**

**Abstract** This chapter discusses the patterns of coupling coefficients between stripline resonators in combline and pseudocombline structures. It was established that there is an electromagnetic coupling between λ/4 (λ/2) resonators in such structures, and they are bandpass filters (BPFs). Between stripline stepped-impedance resonators (SIRs), both positive and negative mixed coupling can be realized. This coupling can vary widely by changing the shape of resonators and the gap between them. Moreover, the tuning of coupling is carried out without the use of a conducting pin, as in the case of microstrip resonators. The patterns of changes in the coupling coefficients between stripline SIRs at higher resonant frequencies were studied. These changes have a wave-like (alternating) character. The effect of transitioning the coupling coefficient through zero can be used to expand the rejection band of BPF by suppressing the nearest spurious bandwidth. It was found that the coupling coefficients between stripline resonators, all of whose side surfaces are metallized, depend only on the geometric parameters and are not dependent on dielectric constant $\varepsilon_r$. The dielectric constant only moves the coupling frequencies of an insulated stripline structure, while maintaining the ratio between these frequencies. The measurement data for some combline and pseudocombline stripline BPFs are presented.

**Keywords** Combline and pseudocombline coupling structures · Discriminating coupling · Invariant of coupling coefficient · Mixed coupling · Rejection band · Transmission zero

A. Zakharov (✉) · M. Ilchenko (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: azakharov217@gmail.com

M. Ilchenko
e-mail: ilch@kpi.ua

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_23

## 1 Introduction

Bandpass filters belong to basic components of microwave telecommunication systems [1, 2]. Recently, they have received considerable attention in the scientific literature [3–8]. Small-sized filters like filters designed on coaxial dielectric resonators [2], monoblock ceramic filters [9], surface acoustic wave (SAW) filters [10], multilayer ceramic filters [11, 12], and microstrip filters [13, 14] are in the great demand. Stripline BPFs, which can overlap a frequency band from 300 to 100 GHz or even higher, show considerable promise [2]. This is due to the possibility of application in these filters of substrates with different values of dielectric constant $\varepsilon_r = 2 - 100$. The use of thick substrates ($h = 2 - 3$ mm) allows us to achieve low losses of these BPF. The use of thin substrates ($h \leq 0.5$ mm) allows us to create thin BPFs with thickness 1 mm or less. Stripline BPF can use a wide variety of metallized topological patterns, making it possible to obtain different frequency responses.

Regardless of this advantage, the development of miniature stripline BPFs in recent decades was very limited. In our opinion, this is related to the proposition about the absence of coupling between quarter-wave ($\lambda/4$) or half-wave ($\lambda/2$) stripline resonators placed in parallel to each other or forming a combline or a pseudocombline structure, which was stated in [15]. This statement entered an authoritative publication on contemporary filters [1] and slowed down the process of development of stripline BPFs.

The combline and pseudocombline stripline structures were attributed to the category of unpromising structures [16]. They were even not included in the review of stripline BPFs presented in 1983 [17]. Only relatively recently it was found [18–20] that there is a coupling between $\lambda/4$ or $\lambda/2$ stripline resonators in combline or pseudocombline structures.

The basis for the design of BPFs is the knowledge of coupling coefficients between resonators and the ability to manage them. The use of mixed coupling coefficients containing magnetic $K_m$ and electric $K_e$ components in BPFs makes their frequency responses more diverse by introducing transmission zeros into them. Such couplings are also called “frequency-dependent” [21] or “resonant” [22]. Mixed couplings are implemented in BPFs on various transmission lines, such as coaxial [23], multilayer [24], microstrip [25–32], andsubstrate integrated waveguide [33–39]. If a BPF has $N$ tuned mixed couplings between adjacent resonators, then we can have $N$ adjustable transmission zeroes at real frequencies. Cited articles differ from each other in ways of implementing mixed coupling coefficients that depend on the transmission line construction, type of resonators used and their parameters. The mixed coupling between resonators of different transmission line construction is studied well. At the same time, the mixed couplings between stripline resonators are not well studied. It can only be noted that this relationship is manifested in closely spaced stepped-impedance resonators [40, 41].

A large number of articles are devoted to the extension of rejection band of BPF. We will select three of them [42–44], which use discriminating couplings between resonators to suppress spurious passbands. In [42] open-loop resonators

were used. In [43] two types of resonators λ/4 and λ/2 were used simultaneously. In [44] discriminating couplings were realized between stepped-impedance resonators (SIRs) of two different types, quarter-wave and half-wave. In these articles microstrip BPFs were considered. It is also advisable to explore the possibility of implementing discriminating couplings between stripline resonators.

This chapter explores the coupling coefficients between resonators in combline and pseudocombline stripline BPFs. The chapter is organized as follows. Section 2 analyzes coupling coefficients between stripline λ/4 resonators. Section 3 analyzes mixed couplings between stripline quarter-wave SIRs. Design of small-sized quasi-elliptic BPF with high dielectric constant $\varepsilon_r = 92$ done. Section 4 analyses coupling coefficients between stripline SIRs of quarter-wave type at higher resonance frequencies. The design of the thin (1 mm) BPF with wide rejection band, achieved due to the zero coupling coefficient at the frequency of the parasitic resonance, is performed. Section 5 considers a new property of electromagnetic interaction between resonators in closed stripline structures. Our conclusion is given in Sect. 6.

## 2 Stripline Quarter-Wave Resonators

Let us consider a pair of stripline quarter-wave resonators (Fig. 1). The stripline design is filled with dielectric having dielectric constant $\varepsilon_r$, its thickness is equal to $b$. Central conductors of these resonators are characterized by width $w$, and length $L$. They are separated by the gap $S$. The bottom and top of the base and the side of the structure are metallized, which makes it insulated. One ends of the central conductors are open and the other one are short-circuited to the conductive wall.

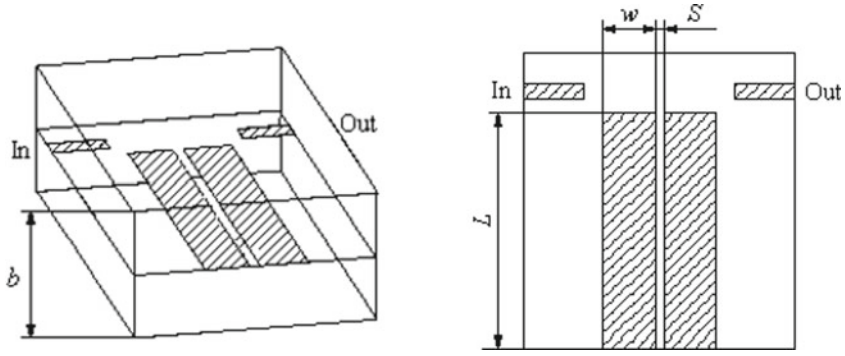



**Fig. 1** Stripline quarter-wave resonator: **a** 3D view; **b** Topology

### 2.1 Coupling Coefficient Between λ/4 Resonators

Each stripline resonator has many resonance frequencies, the lowest of which $f_0$ is the main one. At frequency $f_0$ the resonator in question is a quarter-wave resonator. Let us consider the electromagnetic interaction of the pair of resonators at the main resonance frequency that is characterized by the coupling coefficient $K$. The values of $K$ can be calculated by the following formula [16]

$$K = \frac{f_o^2 - f_e^2}{f_o^2 + f_e^2} \tag{1}$$

where $f_e$ and $f_o$ are the frequencies of even and odd modes of oscillations, respectively. The insertion loss response of a pair of stripline resonators weakly coupled with input and output loads contains two pronounced peaks corresponding to the coupling frequencies. The simulation of frequency responses involved the use of commercial program Microwave Office (AWR Company). Excitation of resonators was performed through capacitive gaps.

Figure 2a illustrates the effect of gap $S$ and dielectric constant $\varepsilon_r$ on coupling coefficient $K$ at the thickness of stripline structure $b = 1$ mm. The simulation was performed by assuming the following values: $\varepsilon_r = 92$, 38, 9.7, and 1; $w = 2$ mm; the value of gap $S$ was varied from 0.05 to 1 mm. At $\varepsilon_r = 92$ the length of resonator was assumed to be equal to $L = 5$ mm, and its corresponding resonance frequency $f_0 \approx 1508$ MHz. At other values of $\varepsilon_r$ the lengths of resonators were selected in such way that their resonance frequencies would coincide with the previous value of $f_0$: at $\varepsilon_r = 38$, $L = 7.9$ mm; at $\varepsilon_r = 9.7$, $L = 15.8$ mm; and at $\varepsilon_r = 1$, $L = 49.5$ mm. The dependencies in Fig. 2a show that the more $\varepsilon_r$, the more $K$.

Figure 2b presents the relationship of $K$ as a function of gap $S$ and dielectric constant $\varepsilon_r$ at doubled thickness of stripline $b = 2$ mm. In this case, the values of $K$ increased more than three times, which indicates that the increase in the thickness of $b$ leads to an increase in $K$.

Figure 2(c) presents the $K$–$S$ relationship for a pair of stripline resonators with $\varepsilon_r = 92$, $b = 2$ mm, $w = 2$ mm, the length of which is reduced to $L = 3$ mm. These resonators have a higher resonance frequency $f_0 \approx 2.3$ GHz. Comparison of dependences in Figs. 2b and c and it shows that the increase of resonant frequency of resonators $f_0$ leads to the growth of $K$. Data in Figs. 2a–c allow us to formulate the patterns of the variation of coupling coefficient $K$ for stripline quarter-wave resonators in combline structure. To increase the $K$ result: increase $\varepsilon_r$; increase the thickness of stripline $b$; increase the resonance frequency $f_0$.

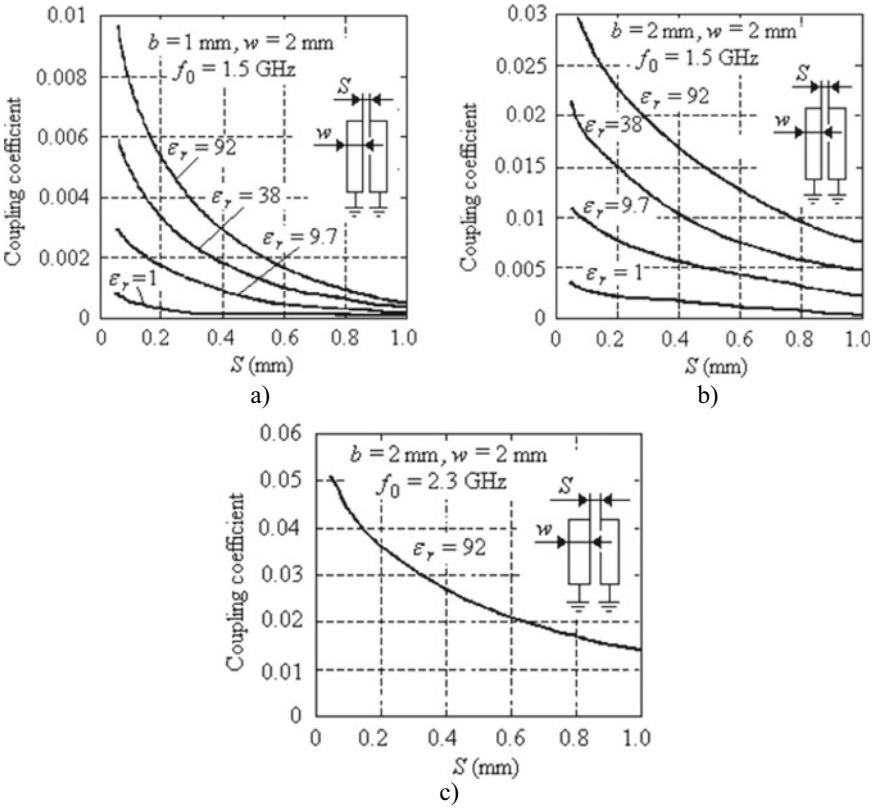



**Fig. 2** Coupling coefficient between stripline λ/4 resonators with wide $w = 2$ mm: **a** for $f_0 = 1.5$ GHz, $b = 1$ mm; **b** for $f_0 = 1.5$ GHz, $b = 2$ mm; **c** for $f_0 = 2.3$ GHz, $b = 2$ mm, $\varepsilon_r = 92$

## 2.2 *Stripline Combline BPF*

Let us design a forth-order stripline combline BPF with Chebyshev response corresponding to the following requirements: center frequency $f_0 = 2380$ MHz; fractional bandwidth FBW = 0.035; passband ripple $L_{Ar} = 0.1$ dB. Using the initial data and parameters of Chebyshev low-pass prototype $g_0$, $g_1$,…, $g_n$, $g_{n+1}$, corresponding to the specified values of $n$ and $L_{Ar}$ and using the known formulas [1]

$$K_{i,\,i+1} = \frac{FBW}{\sqrt{g_i g_{i+1}}} \quad \text{for } i = 1, \ldots, n-1 \tag{2}$$

$$Q_{e1} = \frac{g_0\, g_1}{FBW}, \quad Q_{e2} = \frac{g_n\, g_{n+1}}{FBW} \tag{3}$$

we can determine the coupling coefficients between adjacent resonators $K_{i,\,i+1}$ and external $Q$-factors, $Q_{e1}$ and $Q_{e2}$ of end resonators. For $L_{Ar} = 0.1$ dB and $n = 4$, using

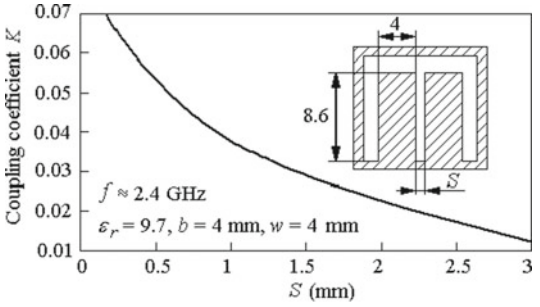
tables [1], we can find $g$-parameters: $g_0 = 1$, $g_1 = 1.1088$, $g_2 = 1.3062$, $g_3 = 1.7704$, $g_4 = 0.8181$, $g_5 = 1.3554$. Substituting these values of $g$-parameters and the initial data into (2) and (3), we obtain $K_{12} = K_{34} = 0.029$; $K_{23} = 0.023$; $Q_e = 31.7$.

The design of the filter is formed by connecting two dielectric substrates from Alumina having thickness $h = 2$ mm, $\varepsilon_r = 9.7$ and $\tan\delta = 0.0002$. The filters fully shielded because all of its side surfaces are metallized.

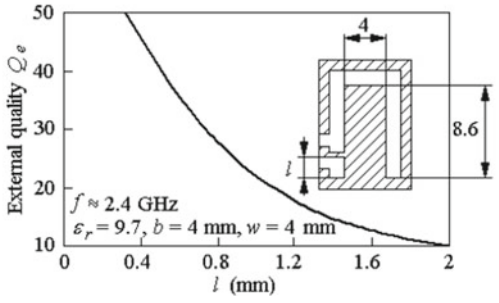
Figure 3 shows the dependencies used in the design, and the insets show filter elements. The central conductors of these resonators are characterized by width $w =$ 4 mm. The short-circuited ends of the resonators are connected to a "protective" band of 1 mm wide, which prevents tin from getting inside the filter during its assembly. The length of resonators, counted from this strip, is equal to $L = 8.6$ mm. The resonators are separated by the gap $S$. The "protective" strip leads to a slight increase in coupling coefficients.

Figure 3a presents the relationship $K = K(S)$ for stripline resonators. According to this curve, we determine that the coupling coefficients $K_{12} = K_{34} = 0{,}029$ and $K_{23} = 0.023$ correspond to the gaps $S_{12} = S_{34} = 1.5$ mm and $S_{23} = 2.0$ mm.

Shown in the inset to Fig. 3b circuit ensures the coupling of end resonator of with load 50 Ω. It represents a segment of stepped-impedance stripline with width 1 and




a)




b)

**Fig. 3** Dependencies used in the design of the stripline BPF: **a** Coupling coefficient; **b** External quality factor

0.5 mm. This segment is removed from the "protective" band by distance *l*. The value of external Q-factor of end resonator $Q_e$ is adjusted by distance *l*.

The values of external *Q*-factor $Q_e$ can be expediently determined by using the results of EM simulation of group delay of parameter $S_{11}$ for resonator with one-sided load [16]. The $Q_e$ can be calculated by using the following expression:

$$Q_e = \frac{\pi \, f_0 \, \tau_{S_{11}}(f_0)}{2} \tag{4}$$

where $\tau_{S_{11}}(f_0)$ is the group delay at the resonance frequency $f_0$. Figure 3b presents the relationship $Q_e = Q_e(l)$ based of (4). The value of external *Q*-factor $Q_e = 31.7$ is obtained at $l = 0.7$ mm.

Dielectric substrates with Alumina were coated with a layer of copper 8 μm thick using vacuum deposition. The topology of central conductors of the filter is shown in Fig. 4a (position *1*), where dimensions are specified in mm. The size of the filter is 26 × 12 × 4 mm. When assembling a filter substrates with conductive layers are pressed to each other according to the "face to face" principle and their metallized ends are soldered together. Conductive strips on the two ends of the filter, which are shown in Fig. 4a (position 2), connect the inner and outer parts of the filter. The latter are contact pads. They are shown in the filter photo presented in Fig. 4b, where the filter is shown on the reverse side. The filter design is intended for surface mounting.

Frequency responses of the stripline BPF are presented in Fig. 5. The measured filter characteristics: center frequency $f_0 = 2380$ MHz, bandwidth BW = 85 MHz, insertion loss at the center frequency $IL_0 = 1.3$ dB, return loss $RL \leq -13.0$ dB, selectivity 40 dB ($f_0 \pm 115$ MHz).

Thus, there is electromagnetic coupling between the stripline combline (λ/4) resonators, and the stripline combline structure is BPF.

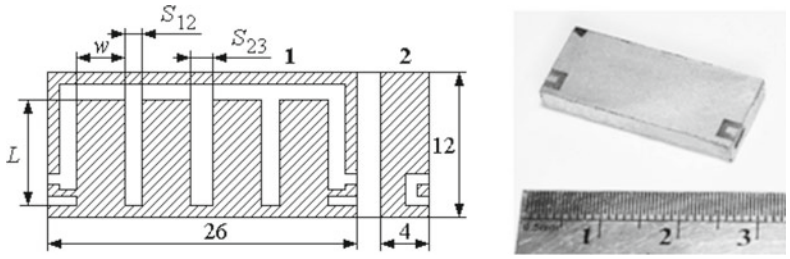



**Fig. 4** Stripline combline BPF: **a** Topology; **b** Photograph of fabricated filter

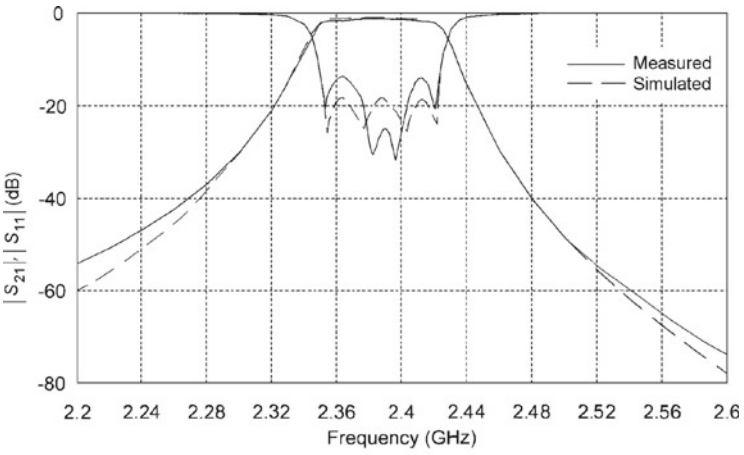



**Fig. 5** Frequency responses of stripline combline BPF

## 3 Mixed Couplings Between Stripline Resonators

The mixed coupling coefficient $K$ has magnetic $K_m$ and electric $K_e$ components [16]

$$K = K_m + K_e. \tag{5}$$

The "+" sign is assigned to the magnetic component and the "–" sign is assigned to the electrical one. In second order BPF with mixed coupling, a $f_z$ transmission zero is generated, the position of which is expressed by the well-known formula [23, 26]

$$f_z = f_0 \sqrt{K_m / |K_e|}. \tag{6}$$

A transmission zero may be above or below $f_0$. If $K > 0$, then $f_z > f_0$, and if $K < 0$, then $f_z < f_0$. The pattern (6) allows us to move the transmission zero $f_z$ with respect to $f_0$ for a fixed value of $|K|$. If the BPF contains $N$ mixed couplings, it can have $N$ transmission zeros.

The design principle of a BPF with mixed couplings between adjacent resonators is as follows. Using expressions (2), the absolute values $|K|$ of mixed coupling coefficients are determined. Each coupling coefficient is associated with a predetermined transmission zero $f_z$. Based on (5), (6), the components $K_m$ and $K_e$ of each mixed coupling coefficient are determined. It is reasonable to study the mixed coupling between stripline SIRs for implementing BPF with pre-defined transmission zeros.

### 3.1 Electromagnetic Interaction Between Stripline SIRs

Figure 6 shows the dependence of $K$ on the form of closely placed stripline SIRs, one end of which is short-circuited. These dependences are constructed for two characteristic shapes of the resonators, which are shown in the insets. The following parameters of SIR were used: the thickness $b = 2$ mm; the dielectric constant $\varepsilon_r = 92$; the resonator length $L = 3.6$ mm; the width of the wide part of the SIR $w_1 = 1.4$ mm; narrow width of the SIR $w_2 = 0.9$ mm. The length $l$ of the narrow section of the resonator is variable. The solid lines correspond to the gap 0.2 mm between the resonators and dashed lines correspond to the gap 0.4 mm. For pair of resonators shown in Fig. 6a, the values of $K$ can be positive, negative, and equal to zero. For positive mixed coupling coefficient ($K > 0$) the magnetic component $K_m$ dominates in it. For $K < 0$ the electrical component dominates. With increasing length $l$ of a narrow part of the resonator (Fig. 6a) the value of $K$ decreases to zero firstly. Then it becomes negative and increases in absolute value. The value of $K < 0$ for almost all values of $l$ is changed. The functions $K = K(l/L)$ are bell-shaped and they have a minimum at $l/L \approx 1/2$. The largest change in $K$ occurs when the gap between the resonators is equal to 0.2 mm: $-0.106 \leq K \leq 0.0385$.

For pair of resonators shown in Fig. 6b, the values of $K$ are always positive. The curves of $K = K(l/L)$ are bell-shaped. They have the maximum at $l/L \approx 1/2$. When the gap between the resonators is 0.2 mm, the values of $K$ vary within $0.0385 \leq K \leq 0.0948$. In the two cases considered, the mixed coupling coefficient between stripline SIRs changed in a wide enough range $-0.106 \leq K \leq 0.0948$. Note that the change in $K$ over a wide range is carried out without the use of an additional conductive pin between two resonators, as in the case of microstrip SIRs [27].

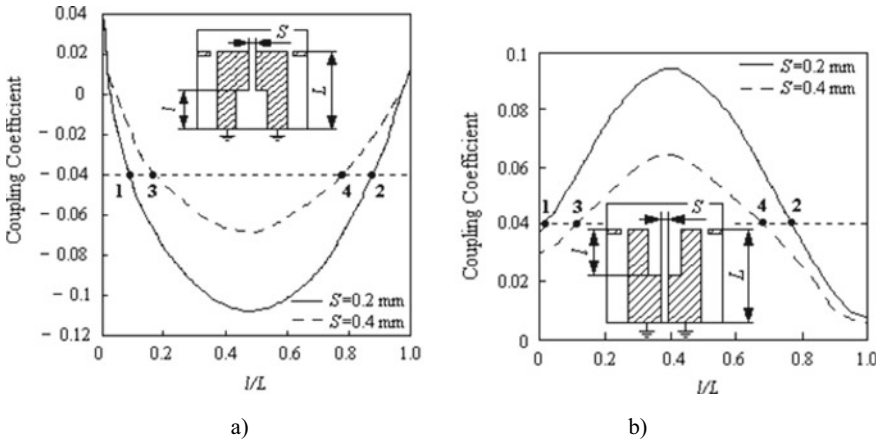



**Fig. 6** Coupling coefficient between stripline SIRs: **a** Short-circuited end of SIR is narrow; **b** Open end of SIR is narrow

The dependences $K(l/L)$, similar to shown in Fig. 6a, can also be obtained if the narrow part of the SIR is located near the open end but in its "inner" part. The dependences $K(l/L)$, similar to shown in Fig. 6b, can also be obtained if the narrow part of the SIR is located near the short-circuited end, but in its "internal" part.

Figure 7 shows symmetric forth-order BPFs with alternating signs of mixed coupling coefficients that have two both sided transmission zeros. Filters in Figs. 7a, b have two couplings with $K < 0$ and one coupling with $K > 0$. The filters in Figs. 7c, d have two positive mixed couplings and one negative mixed coupling.

The coupling coefficients $K$ in Figs. 6a, b are mixed. Expressions (5), (6) allow us to determine the components of $K_{\mathrm{m}}$ and $K_{\mathrm{e}}$ if the values of $K$ and $f_{\mathrm{z}}$ are known. The center frequency $f_0$ in (6) is defined as follows: $f_0 = (f_{\mathrm{o}} + f_{\mathrm{e}})/2$. Table 1 shows the values of $K_{\mathrm{m}}$ and $K_{\mathrm{e}}$ for four points in Fig. 6a with the same negative mixed coupling coefficient $K = -0.04$. Each curve in Fig. 6a intersects line $K = -0.04$ at two points. Points 1 and 2 correspond to the gap $S = 0.2$ mm between resonators. Points 3 and 4 correspond to the gap $S = 0.4$ mm. Figure 8a shows the SIRs topologies corresponding to these points.

Table 2 presents the values of $K_m$ and $K_e$ for four points in Fig. 6b with the same positive mixed coupling coefficient $K = 0.04$. As in the previous case, points 1 and 2 correspond to the gap $S = 0.2$ mm between two resonators. Points 3 and 4 correspond

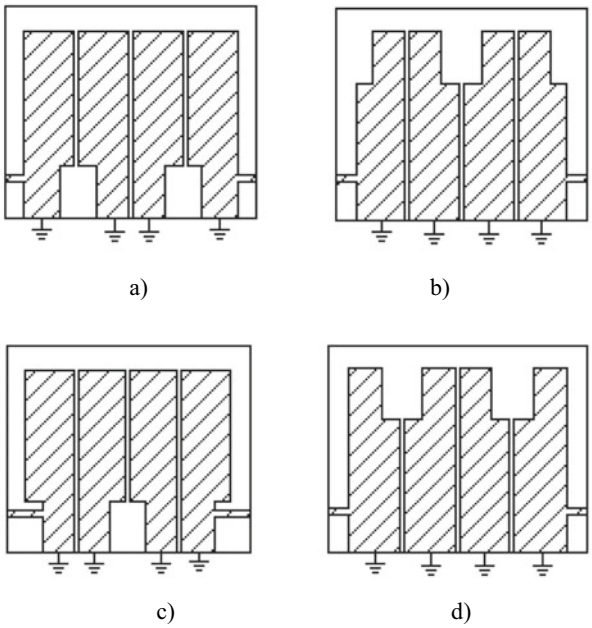
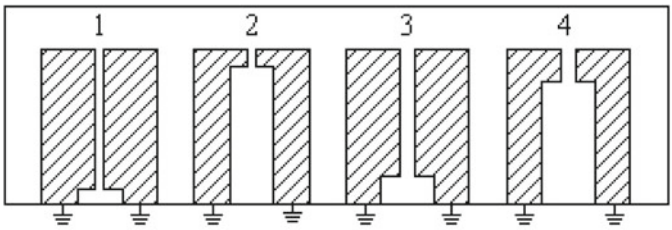


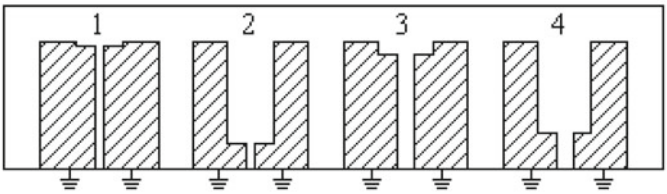

**Fig. 7** Forth-order stripline BPFs with alternating mixed couplings: **a** Two negative and one positive coupling, short-circuited end of SIR is narrow; **b** Two negative and one positive coupling, open-circuited end of SIR is narrow; **c** Two positive and one negative couplings, short-circuited end of SIR is narrow; **d** Two positive and one negative couplings, open-circuited end of SIR is narrow

**Table 1** Components of mixed coupling $K = -0.04$ at points 1, 2, 3, 4 in Fig. 6a

| No. | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $l/L$ | 0.0903 | 0.8819 | 0.1806 | 0.7917 |
| $K_m$ | 0.1295 | 0.0642 | 0.0658 | 0.0479 |
| $K_e$ | −0.1695 | −0.1042 | −0.1058 | −0.0879 |
| $K_m/\|K_e\|$ | 0.764 | 0.616 | 0.622 | 0.545 |




a)




b)

**Fig. 8** Pairs of SIRs with mixed coupling coefficient $|K| = 0.04$: **a** Pairs of SIRs corresponding to points 1, 2, 3, 4 in Fig. 6a for $K = -0.04$; **b** Pairs of SIRs corresponding to points 1, 2, 3, 4 in Fig. 6b for $K = 0.04$

to the gap $S = 0.4$ mm. Figure 8b shows the SIRs topologies corresponding to these points.

The data presented in Tables 1, 2 and Fig. 8 indicate that the mixed coupling coefficient $K$ can be implemented on a set of SIRs of various shapes and with different gaps between resonators. The differences between these pairs of resonators lie in

**Table 2** Components of mixed coupling $K = -0.04$ at points 1, 2, 3, 4 in Fig. 6b

| No. | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $l/L$ | 0.0139 | 0.7917 | 0.0972 | 0.7917 |
| $K_m$ | 0.2305 | 0.0829 | 0.1559 | 0.0699 |
| $K_e$ | −0.1905 | −0.0429 | −0.1159 | −0.0299 |
| $K_m/\|K_e\|$ | 1.210 | 1.932 | 1.345 | 2.338 |

the values of the components $K_m$ and $K_e$, and, therefore, in the position of the transmission zero $f_z$ relative to $f_0$. The shape of the resonators and the gap between them allow us to adjust the position of $f_z$ relative to $f_0$.

Analysis shows the following patterns:

1. The convergence of stripline SIRs, i.e. decreasing the gap $S$ between them leads to an increase of the components $K_m$ and $K_e$ for a fixed value of $K$. In this case the ratio $K_m/|K_e|$ approaches unity and interval between frequencies $f_z$ and $f_0$ decreases.
2. For fixed gap $S$ between stripline SIRs a smaller ratio of lengths $l/L$ leads to a smaller interval between frequencies $f_z$ and $f_0$.
3. The degree of convergence of the frequencies $f_z$ and $f_0$ is limited by the minimum allowable gap $S$ between stripline resonators.
4. Two mixed coupling coefficients $K_1$ and $K_2$ are equal to each other if $K_{1m} = K_{2m}$ and $K_{1e} = K_{2e}$. The mixed coupling coefficients at points 1, 2, 3, 4 (Fig. 6) do not satisfy these conditions.

### 3.2 Small Sized Stripline Quasi-Elliptic BPF

We realize a fourth-order stripline combline filter. This filter has a Chebyshev characteristic in the passband and following specification:

- Center frequency: $f_0 = 1835$ MHz.
- Bandwidth: BW = 90 MHz (FBW = 0.049).
- Return loss: $RL < -13.5$ dB.
- Transmission zeros: $f_{z1}, f_{z2} \in [f_0 \pm 200\ \text{MHz}]$.

In this case two transmission zeroes are not defined exactly. They should be removed from $f_0$ no more than 200 MHz providing increased frequency selectivity. The specified value of return loss $RL = -13.5$ dB corresponds to the ripple value: $L_{\text{Ar}} \approx 0.2$ dB. For designing the filter with the values $L_{\text{Ar}} \approx 0.2$ dB and $n = 4$ the next $g$-parameters are used [1]: $g_0 = 1$, $g_1 = 1.3028$, $g_2 = 1.2844$, $g_3 = 1.9761$, $g_4 = 0.8468$, $g_5 = 1.5386$. Substituting these values and the value $FBW = 0.049$ into (2) and (3) we obtain $|K_{12}| = |K_{34}| = 0.0379$, $|K_{23}| = 0.0308$, $Q_e = 26.59$. We take the values of $K_{12}$, $K_{34}$ positive and $K_{23} < 0$.

These coupling coefficients are mixed and their components $K_m, K_e$ must provide the required location of the transmission zeros. The transmission zero for the positive $K_{12}$ value is denoted by $f_{z1}$. According to the accepted specification we have $f_{z1} \leq f_0 + 200$ MHz. From this inequality, we define $f_{z1}/f_0 \leq 1.109$. Using (6) we write the condition that the components $K_{m12}$ and $K_{e12}$ must satisfy:

$$K_{m12}/|K_{e12}| \leq 1.23. \tag{7}$$

The transmission zero for a negative $K_{23}$ value is denoted as $f_{z2}$. According to the accepted specification we have $f_{z2} \geq f_0 - 200$ MHz, whence it follows $f_{z2}/f_0 \geq$

0.891. Using (6) we finally get

$$K_{m23}/|K_{e23}| \geq 0.794. \tag{8}$$

In the developed stripline BPF two thermostable dielectric substrates with $\varepsilon_r = 92$, $\tan\delta = 0.0003$ are used. The thickness of the substrates is $h = 1$ mm and thickness of the stripline filter is $b = 2$ mm.

Figures 9a, b show the topology of the resonator pairs of the developed filter. Both pairs contain the same resonators, which are oriented relative to each other in different ways. As a result, the pair of resonators in Fig. 9a is characterized by the value $K < 0$ and the pair of resonators in Fig. 9b is characterized by the value $K > 0$.

The short-circuited ends of the resonators are connected to a "protective" strip of 0.4 mmwide. The length of each resonator, counted from this strip, is equal to $L = 3.2$

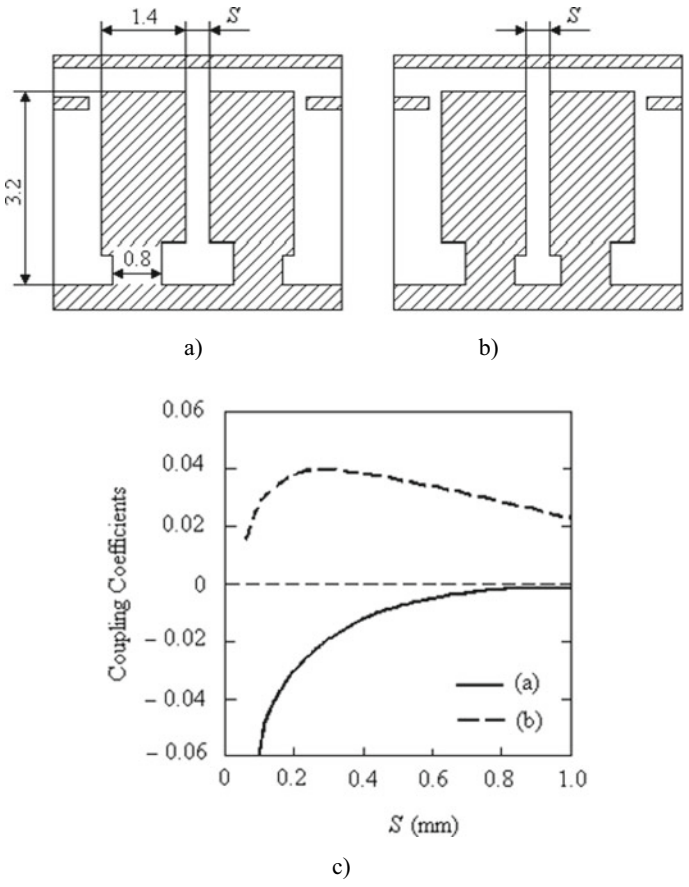



**Fig. 9** Resonator pairs of the developed BPF with mixed coupling coefficients: **a** Pairs of SIRs with $K < 0$; **b** Pairs of SIRs with $K > 0$; **c** Dependence of the value $K$ on the gap $S$

mmand the maximum width of the resonator is $w = 1.4$ mm. In the area of the short circuit two rectangles of 0.8 mm × 0.4 mm and 0.6 mm × 0.2 mm were extracted from the central strip of the resonator. As a result, the resonator becomes step-impedance and tuned to the frequency $f_0$. The protective strip in the short-circuit end region of the resonators increases the magnetic component $K_m$ of the mixed coupling. The wider it is the more $K_m$. This led to the fact that in pair of resonators (Fig. 9b) with $K > 0$, the width of the conductors was reduced in the area of the short circuit ends.

Using computer simulation we present the dependences $K = K(S)$ in Fig. 9c for the considered pairs of resonators. Values $K_{12} = K_{34} = 0.0379$ are realized on a pair of resonators shown in Fig. 9b at $S_{12} = S_{34} = 0.2$ mm. The value of $K_{23} = -0.0308$ is implemented on a pair of resonators shown in Fig. 9a at $S_{23} = 0.2$ mm. Using (5), (6), we determine the components of $K_{m12}$ and $K_{e12}$ of the mixed coupling $K_{12}$: $K_{m12} = 0.2079$; $K_{e12} = -0.170$. These components satisfy the condition (7), which is derived from the specification. Similarly, we determine the components $K_{m23}$ and $K_{e23}$ of the mixed coupling $K_{23}$: $K_{m23} = 0.1547$; $K_{e23} = -0.1855$. These components satisfy the condition (8).

Combining the pairs of resonators shown in Fig. 9 leads to a fourth-order BPF shown in Fig. 10. Dielectric substrates with $\varepsilon_r = 92$ and $\tan\delta = 0.0003$ were coated with a layer of copper 8 μm thick using vacuum deposition. The topology of the central conductors of the filter is shown in Fig. 10a (position 1), where dimensions are indicated in mm. The size of the filter is 7.4 × 4.2 × 2 mm. The use of the "protective" strip of 0.4 mm wide has resulted in a different topology for this filter from the filter topology shown in Fig. 7b. Conductive strips on the two ends of the filter, which are shown in Fig. 10a (position 2), connect the inner and outer parts of the filter. The latter are contact pads. They are shown in the filter photo presented in Fig. 10b, where the filter is shown on the reverse side. The required external quality factor of the end resonators $Q_e = 26.59$ is provided by the appropriate choice (4) of the connection coordinates of the I/O strips.

The measured and simulated frequency responses of the filter are shown in Fig. 10c.

The measured filter characteristics: center frequency $f_0 = 1835$ MHz, BW = 90 MHz, insertion loss at the center frequency $IL_0 = 1.9$ dB, insertion loss at the edges of the passband $IL_{max} = 3.2$ dB, return loss $RL \leq -14.0$ dB.

Two transmissions zeros of the filter are located at frequencies $f_{z1} = 2034$ MHz and $f_{z2} = 1668$ MHz, which correspond to the previously accepted specification. The filter has an increased selectivity of 45 dB ($f_0 \pm 135$ MHz). When using a normalized frequency $j\Omega = (f/f_0 - f_0/f)$/FBW the selectivity can be presented in the form of 45 dB ($\pm j3$). For comparison, we note that the forth-order "High-Perf Ceramic Monoblock Filter CER0206A" [45] with dimensions 7.1 mm × 5.6 mm × 4 mm, $f_0$ = 1880 and BW = 60 MHz, has lower selectivity: 42 dB($-j3.3$); 24 dB($j3.3$).

Extremely important is the effect of the transition of the mixed coupling coefficient through the zero value, as shown in Fig. 6a. In this moment the components of the mixed coupling are equal to each other $K_m = |K_e|$. If a hairpin resonator is placed over a pair of such resonators, and open ends of this hairpin resonator are turned in the direction of the first two resonators, a third-order BPF is formed. The mixed coupling

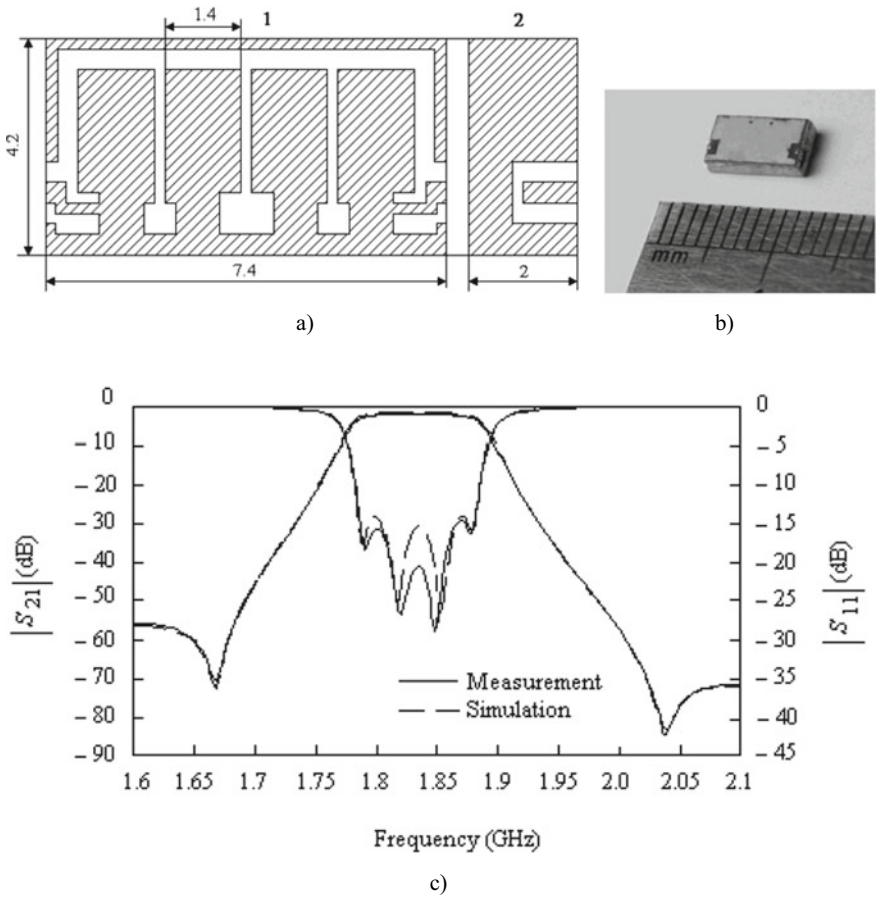



**Fig. 10** Miniaturized quazi-elliptic stripline BPF with mixed coupling: **a** Topology; **b** Photograph of fabricated filter; **c** Frequency responses

of the formed filter is cross- coupling one. Such a filter has a quasi-elliptic frequency response with two transmission zeros located equidistant relative to $f_0$ [28]. The larger the value of $K_m = |K_e|$, the closer the transmission zeros to $f_0$. If the hairpin resonator is replaced with a quarter-wave resonator, then the newly formed filter will have a flat group delay [29]. These effects were demonstrated in [28] for microstrip BPFs. But they can also be used in stripline BPFs that are self-shielded, and their resonators have a higher unloaded quality factor $Q_u$ than microstrip resonators. The transition of the mixed cross-coupling coefficient through the zero value gives the third-order filter the properties that are inherent in a quadruplet BPF with all simple couplings.

The use of substrates with a low dielectric constant ($\varepsilon_r = 2.2$) allows the filters in question to operate at frequencies of 100 GHz and higher, which is consistent with the results given in [2].

## 4 Coupling Coefficients at Higher Resonant Frequencies

Figures 11a, b show two types of low-profile ($b = 1$ mm) stripline SIRs with dielectric constant $\varepsilon_r = 9.7$ (Alumina), whose length is $L = 10$ mm. The maximum width of the resonator is 1.6 mm($Z_{0\min} = 14.75\ \Omega$), and the minimum width of the resonator is 0.8 mm($Z_{0\max} = 24.26\ \Omega$). The resonators are separated by the gap $S = 0.1$ mm.

### *4.1 Coupling Coefficients vs Shape of SIRs*

These resonators are characterized by many oscillations with resonant frequencies $f_n$, $n = 0, 1, \ldots$, which are electromagnetically coupled to each other. Let's denote $K_0$, $K_1$, and $K_2$ coupling coefficients corresponding to resonance frequencies $f_0$, $f_1$ and $f_2$. Figures 11c, d show the dependences of $K_0$, $K_1$, and $K_2$ on the normalized value $l/L$ of these SIRs.

Let us mention the characteristic features of curves $K_n = K_n(l/L)$ shown in Figs. 11c, d:

1. Functions $K_n(l/L)$ demonstrate alternating behavior; the number of extremes of these curves is $2n + 1$.
2. The number of internal zeros of function $K_n(l/L)$ is $2n$. The coupling coefficients $K_0$, $K_1$ and $K_2$ are mixed. Condition for suppression of the first spurious passband $K_1 = 0$ is of the greatest interest, because, in this case, the rejection band of the BPF expands.

Plots in Fig. 11c show that the condition $K_1 = 0$ holds for $l/L \approx 1/3$ and $l/L \approx 2/3$ for both types of resonator pairs. To determine which pair of resonators in Figs. 11a, b give preference, additional analysis is necessary.

Resonance frequencies of considered SIRs (Figs. 11a, b) can be determined from the resonance equations.

$$m \tan\left[\omega(L-l)/v\right] = -\cot\left(\omega l/v\right) - \text{for first type;} \tag{9'}$$

$$m^{-1} \tan\left[\omega l/v\right] = -\cot\left[\omega(L-l)/v\right] - \text{for second type} \tag{9''}$$

where $v$ is the propagation speed of electromagnetic wave, and $m = Z_{0\max}/Z_{0\min}$. Figure 12 shows the plots of the ratio of resonance frequencies $f_2/f_0$, which characterize the distance between these frequencies. These curves show that the SIR of the

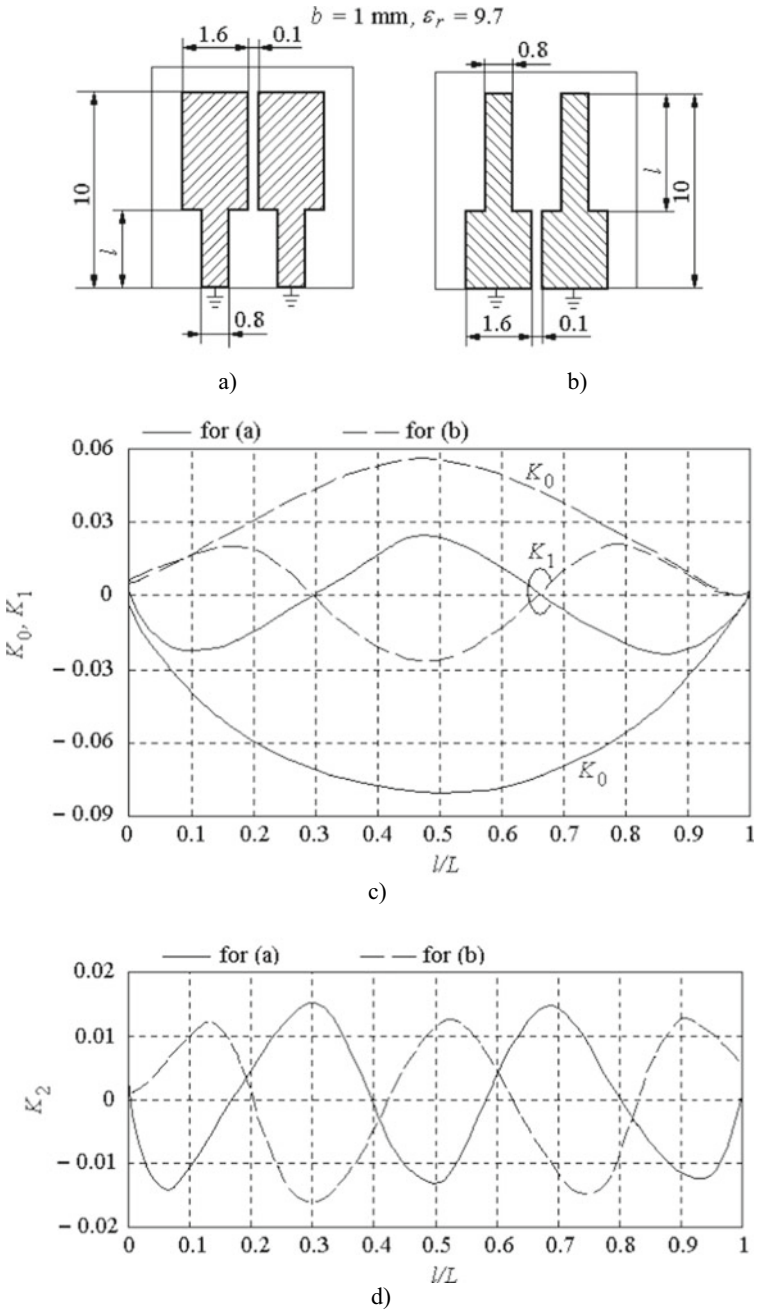



**Fig. 11** Coupling coefficients between low-profile stripline SIRs: **a** First type SIRs; **b** Second type SIRs; **c** Coupling coefficients $K_0$ and $K_1$; **d** Coupling coefficients $K_2$

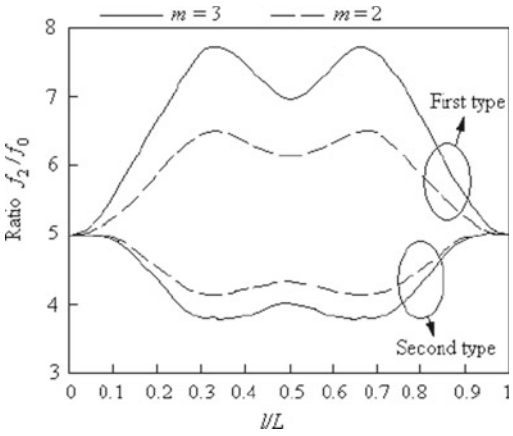



**Fig. 12** Ratio $f_2/f_0$ of stripline SIRs

first type is more preferable because it has higher value $f_2/f_0$. For a pair of resonators of the first type, upper curves $f_2/f_0$ (Fig. 12) have two maximum: at $l/L = 1/3$ and at $l/L = 2/3$. It follows from expression (9′) that, for the SIR, the ratio $f_2/f_0$ is given by the formula

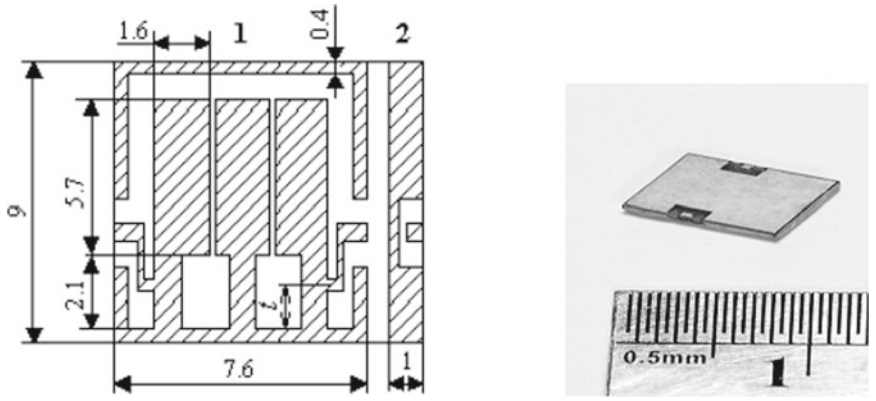
$$\frac{f_2}{f_0} = \frac{\pi}{\tan^{-1}\sqrt{1/(1+2m)}} - 1 \tag{10}$$

from which we find that $f_2/f_0 = 6.47$ for $m = 2$ and $f_2/f_0 = 7.69$ for $m = 3$.

Comparison of the curves in Figs. 11c and 12 shows that there is a double effect at points $l/L \approx 1/3$ and $l/L \approx 2/3$: the zero value of $K_1 = 0$, and the maximum ratio $f_2/f_0$. Combining both effects, we can obtain a significant expansion of the BPF rejection band.

## *4.2 Low Profile Stripline BPF with Extended Stop Band*

To demonstrate the described above effect ($K_1 = 0$), a third-order Chebyshev filter centered at $f_0 = 2.45$ GHz, with a 0.2 dB bandpass ripple and BW = 140 MHz (FBW = 0.057), was designed. According to [1] these values give the coupling coefficients $K_{12} = K_{23} = K_0 = 0.048$ and external quality factor of end resonators $Q_e = 21.5$. The low profile ($b = 1$ mm) stripline BPF with SIRs was fabricated from the Alumina substrates with thickness $h = 0.5$ mm. Topology of the filter is depicted in Fig. 13a, position 1. Some dimensions are in millimeters: $l = 2.1$ mm, $L = 7.8$ mm ($l/L = 0.27$), $w_{\max} = 1.6$ mm ($Z_{0\min} = 14.75\ \Omega$), $w_{\min} = 0.8$ mm ($Z_{0\max} = 24.26\ \Omega$), $S = 0.2$ mm, $t = 1.2$ mm, BPF size is 9 mm × 7.6 mm × 1 mm. Conducting segments




a) b)

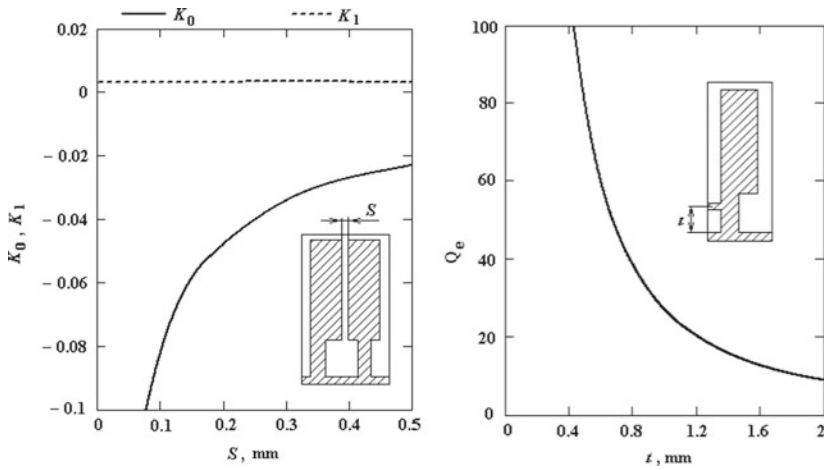
**Fig. 13** Low-profile stripline BPF with $K_1 = 0$: **a** Topology; **b** Photograph of fabricated filter

were formed on the BPF faces (Fig. 13a, position 2). These segments connect internal and external filter parts. The external part is depicted in the filter photo presented in Fig. 13b, where the filter is shown on the reverse side.

Figure 14a plots coupling coefficients $K_0$ and $K_1$ with various gap values $S$. If $S = 0.2$ mm, then $|K_0| = 0.048$. Although the resonators in the inset to Fig. 14a are not identical as in Fig. 11a, however the coupling coefficient $K_1 = 0.0032$ is quite small for given parameters. Figure 14b plots external quality factor $Q_e$ with different $t$ values. If $t = 1.2$ mm, then $Q_e = 21.5$.

The measured and simulated frequency responses of the filter are shown in Fig. 15. The measured date: $f_0 = 2450$ MHz; BW = 140 MHz; $IL_0 = 1.8$ dB; $RL < -16$ dB. Spurious bandpass associated with resonant frequency $f_1$ is suppressed as predicted. First spurious bandpass of the BPF is $6.06\,f_0$. SIRs of the filter are characterized by the parameter $m = 24.26\ \Omega/14.75\ \Omega = 1.645$. Substituting the value of $m$ in (10), we get the calculated value of $f_2/f_0 = 5.98$, which is close enough to the measured value.

The rejection band of this filter at the attenuation level 30 dB is characterized by the ratio of the boundary frequencies $f_{\max}/f_{\min} = 5.44$. For comparison, we note that third order BPF [43] with $\lambda/2$ and $\lambda/4$ resonators, using discriminating coupling $K_1$ has rejection band from 2.1 to 8.6 GHz ($f_{\max}/f_{\min} = 4.09$).




a) b)

**Fig. 14** Design plots for low profile stripline BPF: **a** Coupling coefficients $K_0$ and $K_1$; **b** External quality factor $Q_e$

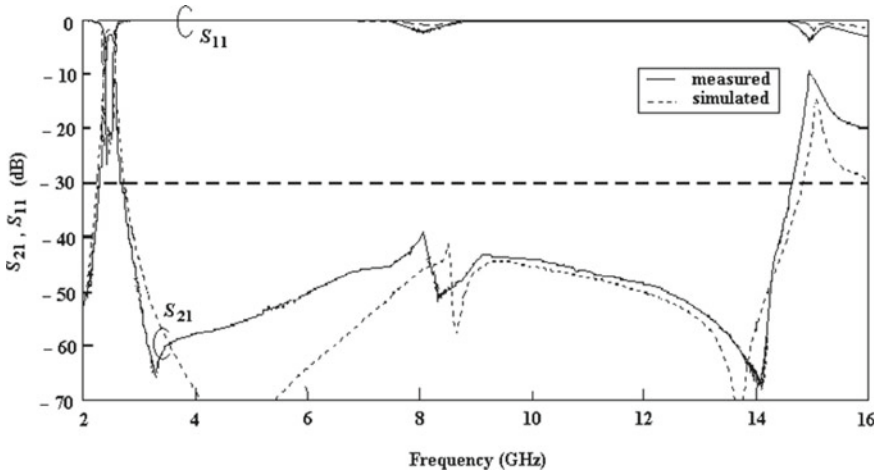



**Fig. 15** Measured and simulated frequency responses of low-profile stripline BPF with $K_1 = 0$

## 5 Invariance of Coupling Coefficients Relative to Dielectric Constant

Consider the influence of the dielectric constant $\varepsilon_r$ on the coupling coefficient $K$ between stripline resonators at the main resonant frequency $f_0$. To exclude the influence of a short circuit on $K$, we will use half-wave resonators.

### 5.1 *Different Location of Stripline Half-Wave Resonators*

Figure 16a shows the topology of half-wave resonators, which are located in the middle of stripline construction. As in Fig. 1a the side surfaces of the strip structure are metallized, which makes the structure insulated. When simulating with Microwave Office (AWR) we use the following parameters: $b = 2$ mm, $L = 6$ mm, $w = 1$ mm, $S = 0.2$ mm. Dimensions of the topological layout are 9.2 mm × 5.4 mm; center conductors are separated from metallized ends by 1.6 mm. In case $\varepsilon_r = 9.7$ we get the following values: $f_e = 7220$ MHz and $f_o = 7604$ MHz. Calculation by (1) gives the value of the coupling coefficient: $K = 0.005177$. If a permittivity $\varepsilon_r = 9.7$ is changed to $\varepsilon_r = 2.2$, then we obtain $f_e = 15159$ MHz, $f_o = 15965$ MHz, and $K = 0.05176$. Table 3 presents the values of $K$ for different values of $\varepsilon_r$. Values of $K$ remain almost unchanged for different values of $\varepsilon_r$, only coupling frequencies $f_e$ and $f_o$ vary.

In Fig. 16b, center conductors of resonators are shifted downward by 0.5 mm. In this case, $K$ varies from 0.0383 to 0.0384 when $\varepsilon_r$ varies from 2.2 to 92. In Fig. 16c, center conductors are shifted downward and to the right, and resonance frequencies of resonators are different. Closeness of metal increases the resonance frequency of the right resonator. If the resonators are no identical, coefficient $K$ is calculated from a more general expression [16]:

$$K = \pm\frac{1}{2}\left(\frac{f_2}{f_1} + \frac{f_1}{f_2}\right)\sqrt{\left(\frac{f_o^2 - f_e^2}{f_o^2 + f_e^2}\right)^2 - \left(\frac{f_2^2 - f_1^2}{f_2^2 + f_1^2}\right)^2}. \tag{11}$$

In (11) $f_1$ and $f_2$ are the resonance frequencies of the first and second resonators. For the pair of resonators shown in Fig. 16c, $K = 0.0352$–$0.0355$ when $\varepsilon_r$ varies from 2.2 to 92. The most general case of arrangement of stripline resonators is shown in Fig. 16d, when center conductors are shifted additionally relative to each other. This pair of resonators has $K = 0.1053$–$0.1055$ when $\varepsilon_r$ varies from 2.2 to 92.

The topology of SIRs of half-wave type is shown in Fig. 16d. Rectangles with dimensions of 1 mm × 0.4 mm are removed in the center part of strips with dimensions of 6 mm × 1 mm. In this case the coupling coefficient is mixed and negative. When $\varepsilon_r$ changes from 2.2 to 92, the value of $K$ changes from –0.0613 to –0.0614.

The results of computer simulation considered above led to the following conclusion. The electromagnetic coupling coefficient of resonators in a stripline structure with homogeneous dielectric filling and metallized lateral surfaces depends only on geometric parameters of the structure and is independent of the value of dielectric constant $\varepsilon_r$.

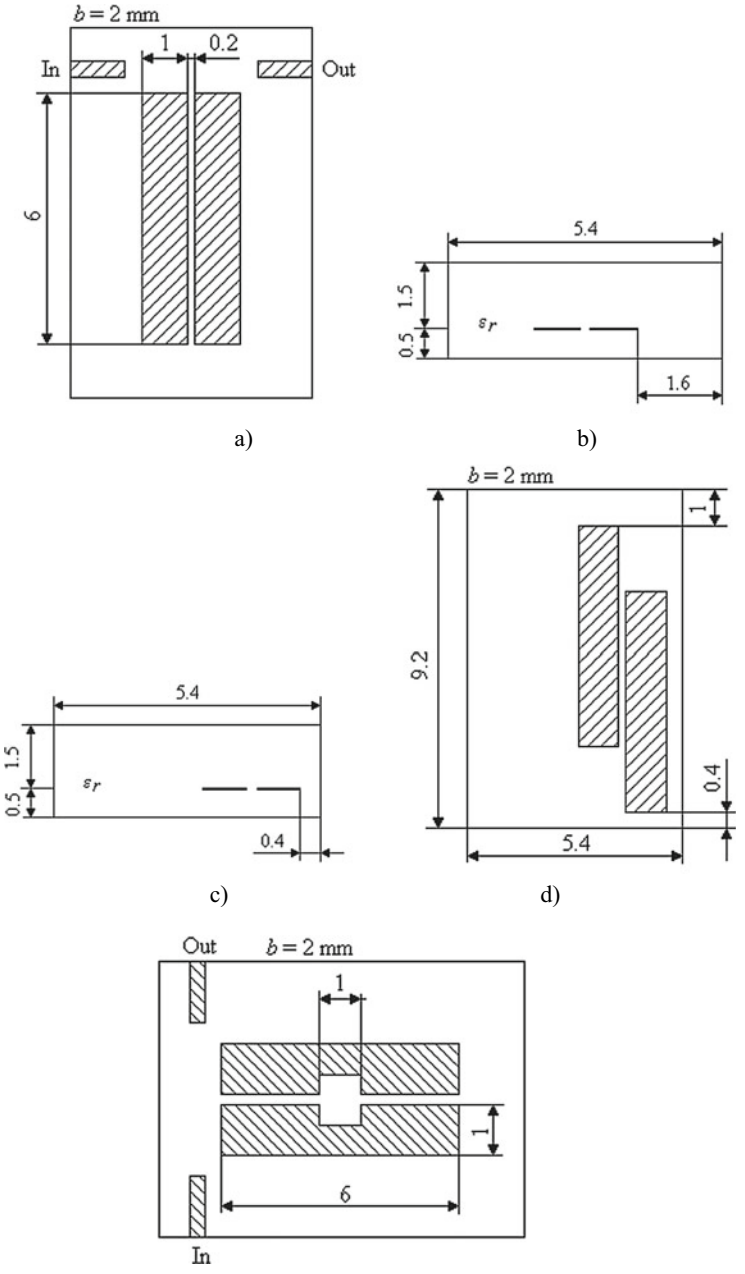



**Fig. 16** Pairs of stripline half wave resonators: **a** Symmetrical positioned; **b** Offset down; **c** Offset down and to the right; **d** Displaced downward, to the right and with respect to each other; **e** SIRs of a half-wave type

**Table 3** Coupling coefficients of stripline resonators (Fig. 16a)

| $\varepsilon_r$ | 2.2 | 9.7 | 21 | 38 | 92 |
|---|---|---|---|---|---|
| $K_{cb}$ | 0.05176 | 0.05177 | 0.05178 | 0.05177 | 0.05170 |

## 5.2 *Dependence of Coupling Coefficient on Resonators' Length*

The influence of geometric parameters, permittivity $\varepsilon_r$, and operating frequency $f_0$ on the value of coupling coefficient $K$ in stripline resonators was analyzed in the Sect. 2. The regularity revealed above leads to the necessity of reconsideration of the influence of last two factors on $K$.

Let us consider the influence of resonator's length on $K$ in more detail. We return to the pair of half-wave stripline resonators (Fig. 16a). Since the permittivity does not affect the value of coefficient $K$, we can use any value of $\varepsilon_r$ in simulation. As before, let us specify remaining parameters of the pair of resonators: $b = 2$ mm, $w = 1$ mm, and $S = 0.2$ mm. The dependence of coupling coefficient $K$ on length $L$ of half- wave stripline resonators is presented in Fig. 17. This dependence is evidence of the strong influence of length $L$ on the coupling coefficient: the smaller $L$, the larger $K$.

It should be noted that quarter-wave resonators have the same coupling coefficients as half-wave resonators, but their length is only half as large. The dependence $K = K(L)$ (Fig. 17) is of practical importance. It is applicable to stripline structures with a great number of permittivity $\varepsilon_r$ values, therefore, it is universal.

Let's return to the dependencies in Fig. 2. It seems that an increase in $\varepsilon_r$ and operating frequency leads to an increase in $K$. Actually, the increase in $K$ of stripline

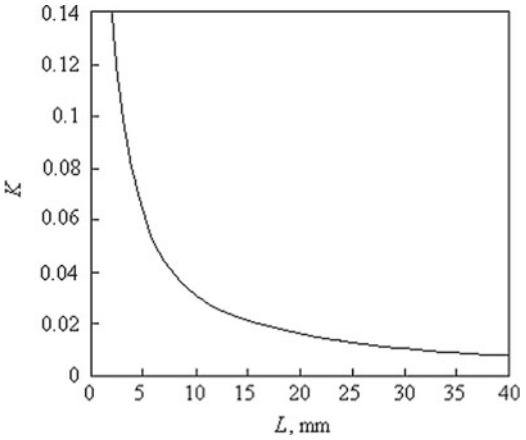



**Fig. 17** Dependence of coupling coefficient on length of stripline half-wave resonators with parameters $b = 2$ mm, $w = 1$ mm, $S = 0.2$ mm

resonators associated with an increase in $\varepsilon_r$ at a fixed frequency as well as with an increase in the operating frequency is caused by a decrease in the resonator length.

The value of $K$ of stripline resonators can also be increased by increasing stripline thickness $b$, decreasing width $w$ of resonator conductor, and decreasing gap $S$ between them. These quantities are geometric parameters of the stripline structure.

### 5.3 Stripline BPF with Different $\varepsilon_r$

The role of the dielectric constant $\varepsilon_r$ is to move coupling frequencies of a stripline without changing the ratio between these frequencies. Let us consider a thin ($b =$ 1 mm) forth-order stripline filter with homogeneous dielectric weakly coupled with input and output loads (Fig. 18a). The filter contains half-wave SIRs. The resonators are characterized by the following parameters: $L = 4.4$ mm and $w = 1.1$ mm. The rectangles with dimensions of 0.9 mm $\times$ 0.4 mm are removed in their center parts. The gap between resonators *2–3* is 0.25 mm, and the gap between resonators *1–2* and *3–4* is 0.3 mm. The filter dimensions are 7.05 mm $\times$ 5.9 mm $\times$ 1 mm. The coupling coefficients of the filter are mixed, and have positive and negative signs.

If the end resonators are weakly coupled with loads, coupling frequencies are clearly seen in the filter's insertion loss function. We denote these frequencies by $f_1, f_2, f_3$ and $f_4$. The coupling frequencies of the filter with dielectric constant $\varepsilon_r =$ 92 are shown in Fig. 18b: $f_1 = 2923$ MHz, $f_2 = 2982$ MHz, $f_3 = 3068$ MHz, and $f_4 = 3105$ MHz. If permittivity $\varepsilon_r$ is decreased by a factor of four and is assumed to be 23, all coupling frequencies increase exactly by a factor of two: $f_1 =$ 5846 MHz, $f_2 = 5964$ MHz, $f_3 = 6136$ MHz, and $f_4 = 6210$ MHz (Fig. 18c). The relation between coupling frequencies in a high-order stripline BPF with homogeneous dielectric and metallized lateral surfaces depends only on geometric parameters and it is independent of the $\varepsilon_r$ value.

This regularity allows us to repeat frequency characteristics of the same filter design in different frequency ranges by changing the value of $\varepsilon_r$. The presented statements highlight that lateral surfaces of the stripline structure should be metallized. In other words, these structures should be insulated electromagnetic structures. Extremely small coupling of end resonators with loads serves as an approximation to such systems. The value of dielectric constant $\varepsilon_r$ does not affect the relation between the coupling frequencies only in the case of complete insulation. When passing to the filter characteristics, it is necessary to increase coupling of the end resonators with loads, the electromagnetic system ceases to be insulated, and frequencies of the end resonated are detuned. The following question arises: what is required for keeping the filter frequency response under variations in $\varepsilon_r$?

Coupling of the end resonators with the load is characterized by external $Q$ factor $Q_e$. The less $Q_e$ the more this coupling. Let us consider the most widespread method for connection of the load to the half-wave resonator via a short conductive section. If we neglect the length of this section, the value of $Q_e$ can be calculated from the expression

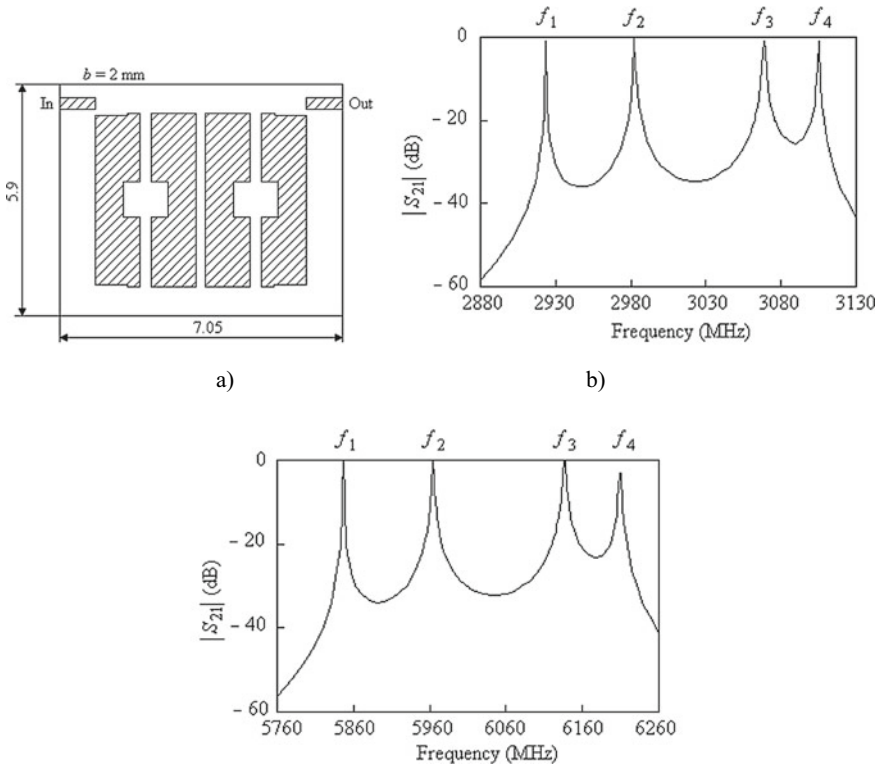



**Fig. 18** Stripline forth order BPF: **a** Topology; **b** Coupling frequencies for $\varepsilon_r = 92$; **c** Coupling frequencies for $\varepsilon_r = 23$

$$Q_e = \frac{\pi R_L}{2 Z_0 \sin^2 \theta'}, \quad \theta' \le \pi / 2 \tag{12}$$

where $R_L = 50\ \Omega$ is the load resistance, $\theta'$ is the load connection coordinate counted from the middle of the resonator. If we reduce the dielectric constant $\varepsilon_r$, then the value of $Z_0$ will increase, and $Q_e$ (12) will decrease. In order to keep the former value of $Q_e$, it is necessary to decrease $\theta'$.

The considered filter (Fig. 18a) contains SIRs, for which expression (12) is not legitimate. However, the pattern of variation in the load connection coordinate associated with the changing of the $\varepsilon_r$ value is retained. Two variants of load connection and frequency responses of the filter, which correspond to values of $\varepsilon_r = 92$ and 23 are shown in Fig. 19. It is important to note that, simultaneously with shifting the load connection point toward the middle point of the resonator, it is necessary to elongate the resonator itself. This neutralizes the frequency drift associated with variation in the load connection coordinate.

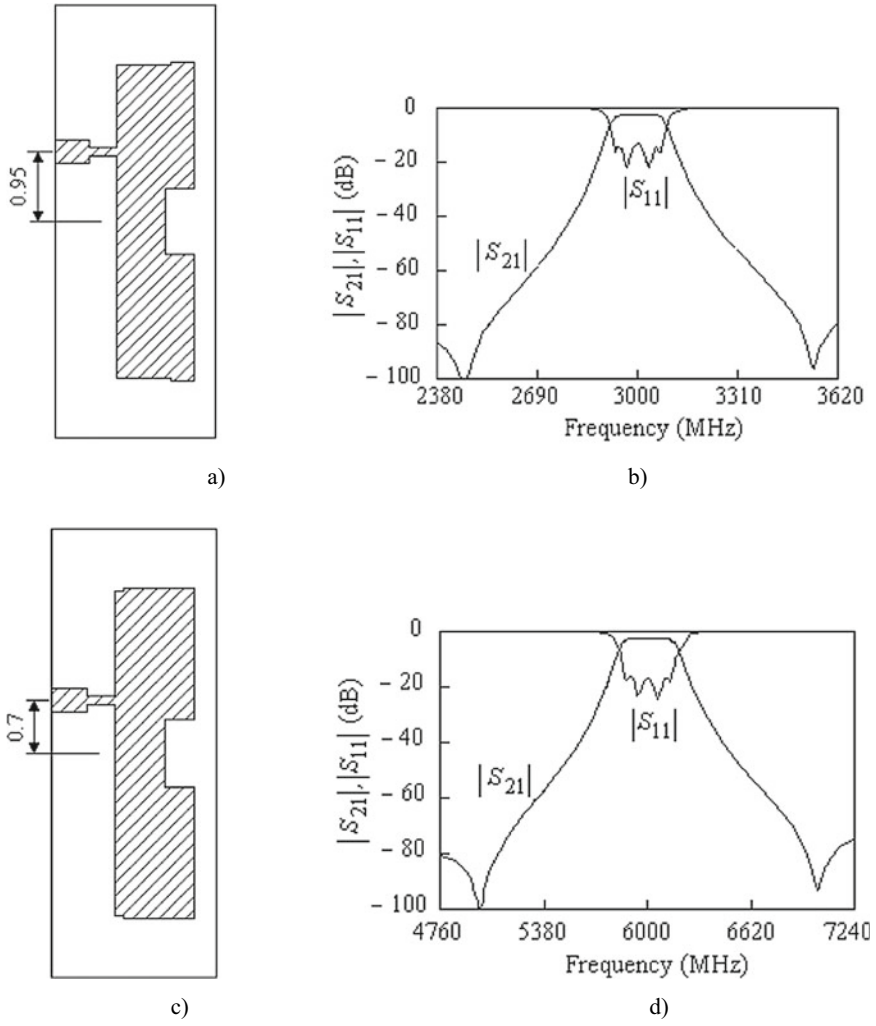



**Fig. 19** Connecting loads and simulated frequency responses of stripline BPF shown at Fig. 18: **a** For case $\varepsilon_r = 92, f_0 = 3000$ MHz, BW = 140 MHz; **b** For case $\varepsilon_r = 23, f_0 = 6000$ MHz, BW = 280 MHz

After changes of $\varepsilon_r$ by a factor of four from 92 to 23, the normalized frequency characteristics remained the same.

The center frequency was doubled from 3000 to 6000 MHz. The absolute bandwidth also changed by a factor of two from 140 to 280 MHz. In general, these changes occur at $\sqrt{\varepsilon_{r2} / \varepsilon_{r1}}$ times.

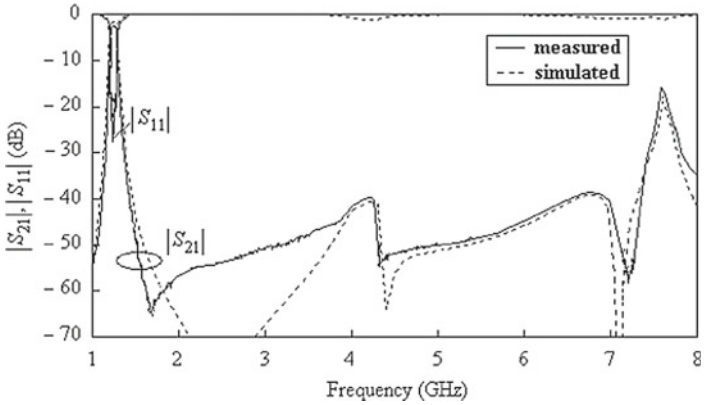



**Fig. 20** Frequency responses of modified low profile stripline BPF, shown in Fig. 13

Thus, if the completed design of a stripline BPF has good frequency response, it can be reproduced in another frequency band by means of appropriate selection of $\varepsilon_r$ and minor changes in the end resonators.

This pattern also applies to a stripline BPF with quarter-wave resonators, as well as to coupling coefficients at higher resonant frequencies. Figure 20 shows the simulated and measured frequency responses of low profile BPF (Fig. 13), whose dielectric constant $\varepsilon_r$ was changed from 9.7 to 38, the end resonators are somewhat shortened, and the coordinate of the connection of loads $t$ is increased from 1.2 to 1.8 mm.

In this case we have $\sqrt{\varepsilon_{r2}/\varepsilon_{r1}} \approx 1.98$. The value of $f_0$ decreased by the same amount from 2450 to 1240 MHz, and BW changed from 140 to 71 MHz. It is important to note that the "zero" value of the coupling coefficient at the frequency of the first spurious resonance $K_1 \approx 0$ is preserved and we get a BPF with a wide rejection band in this case as well.

Note that when forming a conductive BPF pattern by etching under the influence of a magnetic field, it is necessary to take into account the effects presented in [46].

## 6 Conclusion

The studied patterns of coupling coefficients in combline and pseudocombline stripline BPF allow us to realize a significant variety of frequency responses. It has been found that there is electromagnetic coupling between the stripline combline ($\lambda/4$) resonators, and the stripline combline structure is BPF. Mixed coupling coefficients between SIRs and the simultaneous use of substrates with high dielectric constant allow us to implement the quasi-elliptic small sized stripline BPFs with high selectivity. Zero coupling coefficient ($K_1 = 0$) at the first spurious resonant frequency of SIRs (Fig. 11c) and thin substrates ($h \leq 0.5$ mm) allows us to create

thin ($b \leq 1$ mm) stripline BPFs with extended rejection band. The established property of the invariance of the coupling coefficients with respect to $\varepsilon_r$ in isolated stripline structures allows us to move the same frequency response on the frequency axis by changing $\varepsilon_r$ only. Stripline BPFs have great potential.

## References


1. Matthaei GL, Young L, Jones EMT (1980) Microwave filters, impedance-matching network, and coupling structures. Artech House, Norwood
2. Makimoto M, Yamashita S (2001) Microwave resonators and filters for wireless communication. Theory, design and application. Springer
3. Perodou A, Korniienko A, Scorletti G, Zarudniev M, David J-B, O'Connor I (2021) Frequency design of lossless passive electronic filters: a state-space formulation of the direct synthesis approach. IEEE Trans Circ Syst I Reg Papers 1(68):161–174
4. Zakharov A (2021) Parametric and structural-parametric synthesis of nonuniform transmission line resonators. IEEE Trans Circ Syst I Reg Papers 3(68):1055–1067
5. Zakharov A, Ilchenko M (2020) Circuit function characterizing tunability of resonators. IEEE Trans Circ Syst I Reg Papers 1(67):98–107
6. Zakharov A, Ilchenko M (2020) Unloaded quality factor of transmission line resonators with capacitors. IEEE Trans Circ Syst I Reg Papers 7(67):2204–2215
7. Zolkov E, Weiss R, Cohen E (2020) Analysis and design of n-path band-pass filters with negative base band resistance. IEEE Trans Circ Syst I Reg Papers 7(67):2250–2262
8. Zakharov A, Litvintsev S, Ilchenko M (2020) Transmission line tunable resonators with intersecting resonance regions. IEEE Trans Circ Syst II Exp Briefs 4(67):660–664
9. Fukasawa A (1982) Analysis composition of new microwave filter configuration with inhomogeneous dielectric medium. IEEE Trans Microw Theory Techn 9(MTT-30):1367–1375
10. Morgan D (2010) Surface acoustic wave filters: with applications to electronic communications and signal processing. Academic Press
11. Yeung LK, Wu K-L, Wang YE (2008) Low-temperature cofired ceramic LC filters for RF applications. IEEE Microw Mag 5(9):118–128
12. Ishizaki T, Fujita M, Kagata H, Uwano T, Miyake H (1994) A very small dielectric planar filter for portable telephones. IEEE Trans Microw Theory Techn 11(MTT-42):2017–2022
13. Zakharov A, Rozenko S, Pinchuk L, Litvintsev S (2021) Microstrip quazi-elliptic bandpass filter with two pairs of anti-parallel mixed-coupled SIRs. IEEE Microw Wirel Compon Lett 5(31):433–436
14. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Hairpin resonators in varactor-tuned microstrip bandpass filter. IEEE Trans Circ Syst II Exp Briefs 10(67):1874–1878
15. Bolljahn JT, Matthaei GL (1962) A study of the phase and filter properties of arrays of parallel conductors between ground planes. Proc IRE 50(3):299–311
16. Hong J-S (2011) Microstrip filters for RF/Microwave application, 2nd edn. Wiley, New York
17. Dworsky LN (1983) Stripline filters. An overview. In: 37th annual frequency control symposium. IEEE, Philadelphia, PA USA, pp 387–393
18. Zakharov AV, Ilchenko MY, Pinchuk LS (2015) Coupling coefficient of quarter-wave resonators as a function of parameters of comb stripline filters. Radioelectron Commun Syst 6(58):284–289
19. Zakharov AV, Ilchenko ME (2015) Pseudocombline bandpass filters based on half-wave resonators manufactured from sections of balanced striplines. J Commun Technol Electron 7(60):801–807
20. Zakharov AV (2013) Stripline combline filters on substrates designed on high-permittivity ceramic materials. J Commun Technol Electron 3(58):265–272

21. Szydlowski L, Lamecki A, Mrozowski M (2012) Coupled-resonator filters with frequency-dependent couplings: coupling matrix synthesis. IEEE Microw Wirel Compon Lett 6(22):312–314
22. Levy R (2004) New cascaded trisections with resonant cross-couplings (CTR sections) applied to the design of optimal filters. In: IEEE MTT-S international microwave symposium digest 2:447–450
23. Wang H, Chu Q-X (2009) An inline coaxial quasi-elliptic filter with controllable mixed electric and magnetic coupling. IEEE Trans Microw Theory Tech 3(57):667–673
24. Hoft M, Shimamura T (2010) Design of symmetric trisection filters for compact low-temperature co-fired ceramic realization. IEEE Trans Microw Theory Tech 1(58):165–175
25. Ma K, Ma J-G, Yeo KS, Do MA (2006) A compact size coupling controllable filter with separate electric and magnetic coupling paths. IEEE Trans Microw Theory Tech 3(54):1113–1119
26. Chu Q-X, Wang H (2008) A compact open-loop filter with mixed electric and magnetic coupling. IEEE Trans Microw Theory Tech 2(56):431–439
27. Zhu F, Hong W, Chen J-X, Wu K (2014) Quarter-wavelength stepped-impedance resonator filter with mixed electric and magnetic coupling. IEEE Microw Wirel Compon Lett 2(24):90–92
28. Zakharov A, Rozenko S, Ilchenko M (2018) Two types of trisection bandpass filters with mixed cross-coupling. IEEE Microw Wirel Compon Lett 7(28):585–587
29. Zakharov A, Ilchenko M (2017) Trisection microstrip delay line filter with mixed cross-coupling. IEEE Microw Wirel Compon Lett 12(27):1015–1017
30. Zakharov A, Litvintsev S, Ilchenko M (2019) Trisection bandpass filters with all mixed couplings. IEEE Microw Wirel Compon Lett 9(29):592–594
31. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Trisection bandpass filter with mixed cross-coupling and different paths for signal propagation. IEEE Microw Wirel Compon Lett 1(30):12–15
32. Zakharov A (2021) Transmission zeros of trisection and quadruplet bandpass filters with mixed cross coupling. IEEE Trans Microw Theory Tech 1(69):89–100
33. Shen W, Wu L-S, Sun X-W, Yin W-Y, Mao J-F (2009) Novel substrate integrated waveguide filters with mixed cross coupling (MCC). IEEE Microw Wirel Compon Lett 11(19):701–703
34. Gong K, Hong W, Zhang Y, Chen P, You CJ (2012) Substrate integrated waveguide quasi-elliptic filters with controllable electric and magnetic mixed coupling. IEEE Trans Microw Theory Tech 10(60):3071–3078
35. Szydlowski L, Jedrzejewski A, Mrozowski M (2013) A trisection filter design with negative slope of frequency-dependent crosscoupling implemented in substrate integrated waveguide (SIW). IEEE Microw Wirel Compon Lett 9(23):456–458
36. Chu P, Hong W, Dai L, Tang H, Chen J, Hao Z, Zhu X, Wu K (2014) A planar bandpass filter implemented with a hybrid structure of substrate integrated waveguide and coplanar waveguide. IEEE Trans Microw Theory Tech 2(62):266–274
37. Tamiazzo S, Macchiarella G (2017) Synthesis of cross-coupled filters with frequency-dependent couplings. IEEE Trans Microw Theory Tech 3(65):775–782
38. Zhao P, Wu K (2019) Cascading fundamental building blocks with frequency-dependent couplings in microwave filters. IEEE Trans Microw Theory Tech 4(67):1432–1440
39. Balewski L, Fotyga G, Lamecki A, Mrozowski M, Mul M, Sypek P, Szypulski D (2020) Step on it! Microw Mag 3(21):34–49
40. Zakharov AV, Ilchenko MYe, Pinchuk LS (2014) Coupling coefficients of step-impedance resonators in stripline band-pass filters of array type. Radioelectron Commun Syst 5(57):217–223
41. Zakharov AV, Ilchenko MY, Karnauh VY, Pinchuk LS (2011) Stripline bandpass filters with step-impedance resonators. Radioelectron Commun Syst 3(54):163–169
42. Zhang XY, Xue Q (2009) Harmonic-suppressed bandpass filter based on discriminating coupling. IEEE Microw Wirel Compon Lett 11(19):695–697
43. Li YC, Zhang XY, Xue Q (2010) Bandpass filter using discriminating coupling for extended out-of-band suppression. IEEE Microw Wirel Compon Lett 7(20):369–371

44. Lin SC, Lin YS, Chen CH (2006) Extended-stopband bandpass filter using both half- and quarter-wavelength resonators. IEEE Microw Wirel Compon Lett 1(16):43–45
45. "High-perf ceramic monoblock filter CER0206A data sheet" CTS electronic components, Bloomingdale, IL (2005). https://www.ctscorp.com
46. Ilchenko MYu, Gorobets OYu, Bondar IA, Gaponov AM (2010) Influence of external magnetic field on the etching of a steel ball in an aqueous solution of nitric acid. J Magn Magn Mater 322(14):2075–2080

# Transmission Line Dual-Mode Resonators and Dual-Band Filters

**Sergii Rozenko, Sergii Litvintsev, and Liudmyla Pinchuk**

**Abstract** The chapter provides an optimal synthesis of nonuniform transmission line structural elements with a special arrangement of resonant frequencies, which made it possible to expand the functionality of dual-mode resonators based on them. Optimal synthesis consists in ensuring a given ratio between the resonance frequencies of structural elements with the minimum value of parameter $m = Z_{0\max}/Z_{0\min}$, which is an optimality criterion. The method of parametric synthesis was used for optimization. It leads to stepped-impedance transmission line segments (SILS) with the required arrangement of resonant frequencies and $m = m_{\min}$. Two SILS pairs are synthesized. As result, the new four dual-mode resonators with enhanced functionality are obtained. All dual-mode resonators have an extended stopband between the main dual-mode oscillation and the nearest parasitic oscillation. At $m = 5$, which is the maximum for planar transmission lines, the parasitic bandpass can be located 6.47 times higher than the main bandpass. The dual-mode resonators have reduced dimensions and their length shortening relative to a half-wave resonator can reach 0.535. The resonators allow increasing their operating frequencies. The excess factor relative to the resonant frequency of a half-wave resonator can be 2.44. This resonator is very promising for use in dual-band BPF, since it allows us to change the relative position of two passbands in the range $1.46 \leq f_2/f_1 \leq 6.47$. This chapter presents the results of EM simulation of four microstrip filters and the measurement results of two microstrip filters. The results obtained can be used in triple-mode resonators.

**Keywords** Dual-band filter · Dual-mode oscillations · Resonant equation · Single oscillations · Stepped-impedance transmission line segment · Stopband

S. Rozenko (✉) · S. Litvintsev (✉) · L. Pinchuk (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Peremogy Avenue 37, Kyiv 03056, Ukraine
e-mail: serge.rozenko@gmail.com

S. Litvintsev
e-mail: litvintsev.sergii@lll.kpi.ua

L. Pinchuk
e-mail: svetovna@email.ua

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_24

## 1 Introduction

Bandpass filters (BPFs) play an important role in modern communication systems and receive significant attention in the present scientific literature [1–7]. BPFs contain resonators made from transmission line sections of various constructions, including microstrip [8–11] and stripline [12–16]. A special place is occupied by dual-mode resonators and BPFs based on them [17–32], since this leads to an increase in a filter selectivity with a smaller number of resonators. Dual-mode resonators have a wide application, they are used in dual-band BPFs [30–32], they are the main cells of a large variety of balanced filters [33], they serve as the basis for the transition to triple-mode resonators [34, 35].

It was first established in [17] that a microstrip ring resonator is a dual-mode resonator. After this article, a significant number of publications have appeared on filters with dual-mode ring resonators [18, 19]. In [20], an dual-mode microstrip resonator was proposed, which is a horizontal half-wave segment with open ends, to the middle of which a quarter-wave open stub is connected. This dual-mode T-shape resonator is used both as a non-tunable dual-mode BPF [21, 22] and as a tunable BPF [23, 24] containing varactors at open-circuited ends.

Almost simultaneously with the resonator [20], another dual-mode resonator was proposed [25]. It has a half-wave resonator with open ends, which is short-circuited in the middle part using a metallized hole. In the resonator circuit, this hole can be represented as a shunt inductance [26] or as a short-circuited stub of short length [27]. If the open ends of this resonator are connected by a capacitor, then a modified loop dual-mode resonator is obtained [28, 29].

The structural elements of these dual-mode resonators are sections of uniform transmission line. These segments are characterized by an equidistant arrangement of resonant frequencies $f_i$, $i = 1, 2, \ldots$, which limits the functionality of dual-mode resonators based on them.

It is known that segments of smoothly [37, 38] and stepped-impedance [39–45] transmission lines allow changing the ratio between their resonant frequencies. A significant legacy is accumulated for stepped-impedance transmission lines [39–45], which is widely used in various applications and allows improving their quality. It is desirable to use this legacy for dual-mode resonators.

This chapter aims to fill this gap. It is assumed that the use of stepped-impedance transmission line segment (SILS) will improve some characteristics of existing dual-mode resonators: to expand the stopband located between the main dual-mode oscillation and the nearest parasitic oscillation; reduce the size; increase their operating frequencies. It is important to be able to vary widely the relative position of the two bandpass in dual-band BPFs. Only in this case, such filters will be functional, since the position of the passbands must correspond to the operating frequency bands of prospective communication systems, and not vice versa. The solution of these problems is associated with a special arrangement of resonant frequencies of SILS.

An important design parameter of SILS is the ratio $m = Z_{0\max}/Z_{0\min}$. For stripline and microstrip transmission lines this parameter is limited to the value of $m \approx 5$.

The limited parameter $m$ leads to the problem of optimal synthesis, in which the required ratio between resonant frequencies must be ensured at the minimum value of $m = m_{\min}$. The article is based on the latest optimization method of resonators [46]. This a parametric synthesis method allows synthesizing SILS with minimum value of $m$. We will use this method to synthesize the structural elements of dual-mode resonators. Note that this method has significant generality. It leads to solutions that are the best among all kinds of smoothly-impedance and stepped-impedance transmission line segments. This generality is reflected in the title of the article. Note that stepped-impedance segments were used in some dual-mode resonators [47] without optimization.

The chapter is organized as follows. Section II solves the problem of synthesizing structural elements for dual-mode resonators with open-circuited ends. The properties of the resonators obtained are studied. When studying these properties, the resonance equations were used, which were established in [48]. Section III synthesizes structural elements for dual-mode resonators with short-circuited ends. The properties of the new resonators are studied on the basis of the resonance equations, which were established in [49]. Traditional even/odd analysis is limited in this case. Section IV presents the results of EM simulation of four microstrip filters and the measurement results of two microstrip filters with new structural elements. Our conclusion is given in Section V.

## 2 Resonators with Open Ends

Consider the symmetric dual-mode resonators [20] and [25] shown in Figs. 1a and 1b respectively. The basis of these resonators is a horizontal half-wave transmission line segment with open ends, to the middle of which an open-circuited stub or short-circuited stub of short length are connected.

### *2.1 Resonance Equations*

The resonators considered can be represented by a two-port network with transfer matrix [*ABCD*]. In the absence of loss, the resonance condition is expressed by one equation.

$$Y_{\text{in}} = 1/Z_{11} = C/A = 0, \tag{1}$$

where $Z_{11}$ is the input impedance of resonator. If zeros of $C$ and $A$ do not coincide in resonator, and $A$ does not have poles different from the poles $C$, the resonance Eq. (1) is equivalent to the condition [48]

$$C = 0. \tag{2'}$$

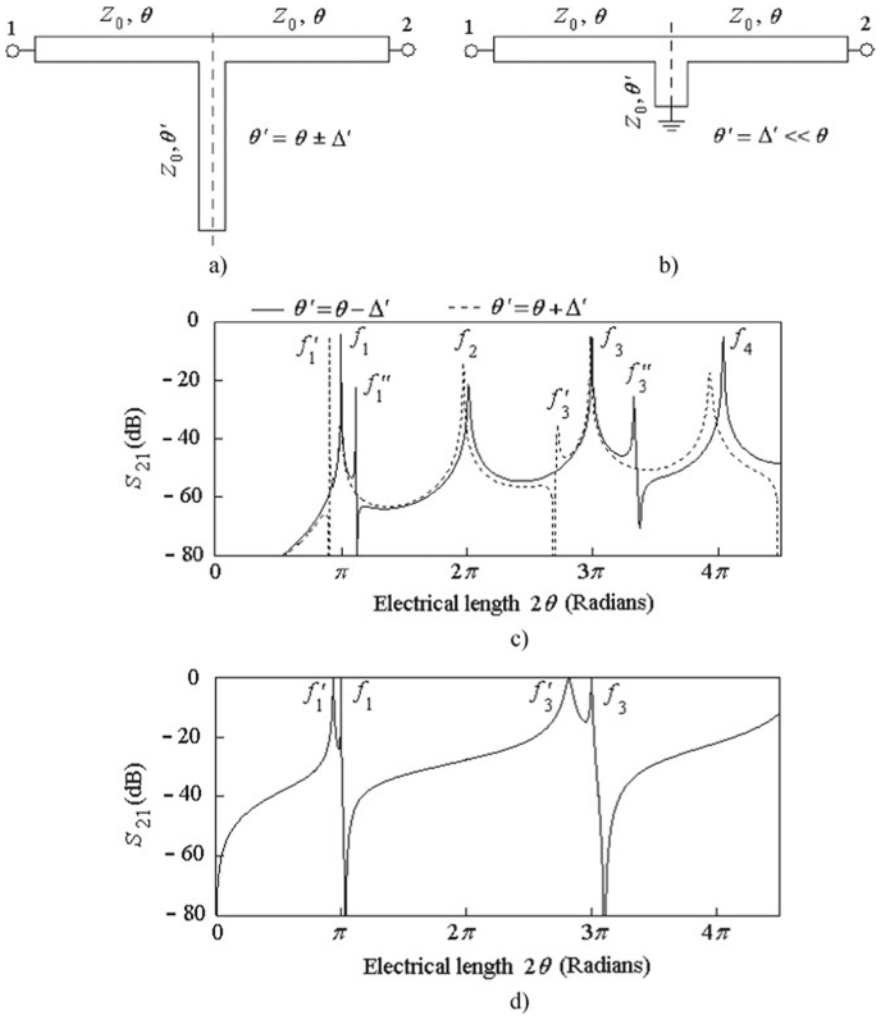



**Fig. 1** Dual-mode resonator with open-circuited ends uniform transmission line segments: **a** topology of resonator with open-circuited stub; **b** topology of resonator with short-circuited stub of short length; **c** resonance frequencies of resonator at Fig. 1a; **d** resonance frequencies of resonator at Fig. 1b

Note that when (1′) is fulfilled, the output conductivity $Y_{\text{out}}$ of the resonator also is equal to zero: $Y_{\text{out}} = 1/Z_{22} = C/D = 0$. The transfer matrix of the symmetrical two-port network [*ABCD*] can be expressed through the transfer matrix elements $A'$, $B'$, $C'$, $D'$ of its left half [50]

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 + 2B'C' & 2A'B' \\ 2C'D' & 1 + 2B'C' \end{bmatrix},$$

which replaces the resonance condition (1′) by two conditions

$$C' = 0 \quad \text{and} \quad D' = 0. \tag{2''}$$

The resonance conditions (1″) are much simpler than the traditionally used condition $Y_{\text{in}} = 0$ (1). They are easier to compose and analyze based on them.

## 2.2 *Properties of Resonators with Uniform Transmission Lines*

Let us compose the resonance equations for the resonators shown in Fig. 1. We will use the input admittance $Y_s$ of the stub at the points of its connection to the transmission line segment, and transfer matrix [*abcd*] of the left half of the transmission line segment. Obviously, the matrix [$A'B'C'D'$] represents the product of two matrices [50]

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ Y_s/2 & 1 \end{bmatrix} = \begin{bmatrix} a + bY_s/2 & b \\ c + dY_s/2 & d \end{bmatrix}. \tag{2}$$

From the last expression it follows that the resonance Eqs. (1″) take the form.

$$d = 0 \quad \text{and} \quad Y_s + 2c/d = 0. \tag{3′), (3”}$$

In general, the matrix [*abcd*] can represent a segment of non-uniform transmission line. If we use a segment of uniform transmission line with characteristic impedance $Z_0$ and electrical length $2\theta$ (Fig. 1), then [51]

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & jZ_0 \sin\theta \\ jZ_0^{-1} \sin\theta & \cos\theta \end{bmatrix}. \tag{4}$$

In this case, resonance Eqs. (3) are follows:

$$\cos\theta = 0 \quad \text{and} \quad Y_s + j2Z_0^{-1} \tan\theta = 0. \tag{5′), (5”}$$

Equation (5′) determines the part of the resonant electrical lengths.

$$\theta_n = \pi/2,\ 3\pi/2,\ 5\pi/2, \ldots. \tag{6}$$

Another part of the resonant electrical lengths is determined by the Eq. (5'') and it depends on the parameters of the used stubs. Let us consider the resonator in Fig. 1a in more detail. Its stub has an electrical length $\theta' = \theta \pm \Delta'$, which differs from $\theta$ by a small fixed amount $\Delta'$, and the input admittance $Y_s = jZ_0{}^{-1} \tan(\theta \pm \Delta')$. The resonance Eq. (5'') for the resonator under consideration takes the form.

$$\tan\left(\theta \pm \Delta'\right) + 2\tan\theta = 0. \tag{7}$$

Equation (7) is transcendental and its exact solutions can be obtained by numerical methods. Equation (7) includes two identical trigonometric functions that are shifted relative to each other along the horizontal axis by a small value $\Delta'$. The approximate solutions of this equation, based on the features of trigonometric functions [52] are presented below. Equation (7) has the solutions in the range of electrical lengths $\theta$ for which the function $\tan\theta = 0$ or $\pm\infty$. In the case $\tan\theta = 0$, the argument $\theta_n = n\pi$ and the approximate equality $\tan(n\pi + \Delta) = \tan\Delta \approx \Delta$ is valid, where $\Delta$ is the small value or detuning. Unlike $\Delta'$, it is a variable. Let $\Delta'$ be a positive value, then equality (7) can be replaced by the approximate.

$$\left(\Delta + \Delta'\right) + 2\Delta \approx 0. \tag{7'}$$

In case $\tan\theta = \pm\infty$, the argument $\theta_n = (2n+1)\pi/2$ and the approximate equality $\tan[(2n+1)\pi/2 + \Delta] = \cot\Delta \approx 1/\Delta$ is valid. Assuming $\Delta'$ to be a positive value, we transform Eq. (7) to the form

$$\frac{1}{\Delta + \Delta'} + \frac{1}{2\Delta} = \frac{3\Delta + \Delta'}{2\Delta(\Delta + \Delta')} \approx 0. \tag{7''}$$

The solution (7') is the meaning $\Delta \approx -\Delta'/3$. It generates the resonant electrical lengths of single oscillations.

$$\theta_n \approx \pi,\ 2\pi,\ 3\pi, \ldots. \tag{8'}$$

Solving (7''), we get the value $\Delta \approx -\Delta'/3$. Returning to the variable $\theta$, we establish that the resonant electrical lengths satisfy the approximate equality.

$$\theta_n \approx \pi/2,\ 3\pi/2,\ 5\pi/2, \ldots. \tag{8''}$$

If the length of the stub is slightly more than a quarter-wavelength, then the resonant electrical lengths are less than the values indicated on the right side (8''), but if the length of the stub is slightly less than a quarter-wavelength, then the resonant electrical lengths are slightly greater than the values (8''). The resonant electrical lengths (6) and (8'') are located next to each other, and they form dual-mode oscillations.

Figure 1c shows the insertion loss frequency responses of this resonator for weak coupling with I/O loads. The solid curve corresponds to a shorter stub ($\theta' = \theta - \Delta$), which has resonant frequencies $f_1$" and $f_3$". If the stub is longer ($\theta' = \theta + \Delta$), which is reflected by the dotted curve, then it characterized by the resonant frequencies $f_1$' and $f_3$'. The resonator under consideration is remarkable in that it has dual-mode oscillations alternating with single oscillations.

If in Fig. 1a the open stub is replaced by a short-circuited stub with short length ($\theta' = \Delta' << \theta$), then we will make the transition to the resonator (Fig. 1b). Its resonant frequencies are shown in Fig. 1d. The part of its resonant electrical lengths is still determined by Eq. (5′) and takes the values (6). In Fig. 1c, they correspond to the resonant frequencies $f_1, f_3, \ldots$. Since the admittance of the stub changed $Y_s = -jZ_0^{-1} \cot\Delta'$, then the resonance Eqs. (5″), (7) take the form.

$$-\cot\Delta' + 2\tan\theta = 0. \tag{9}$$

The resonance Eq. (9) differs significantly from the analogous Eq. (7) of the previous resonator (Fig. 1a). It no longer generates resonant electrical lengths (8′), at which the function $\tan\theta = 0$, and which correspond to simple oscillations with even numbers (Fig. 1c). A short-circuited stub of short length generates extreme left-hand oscillations of dual-mode oscillations with frequencies $f_1'$, $f_3'$, …. A characteristic feature of the considered one (Fig. 1b, 1d) is that all its oscillations are dual-mode.

The resonance Eqs. (5) of the considered resonators (Fig. 1) make it possible to accurately determine their resonant frequencies. This makes it possible to calculate the coupling coefficient $K$ between adjacent dual-mode oscillations [53]

$$K = \left(f_{\text{odd}}^2 - f_{\text{ev}}^2\right)/\left(f_{\text{odd}}^2 + f_{\text{ev}}^2\right), \tag{10}$$

where $f_{\text{odd}}$ and $f_{\text{ev}}$ are the odd and even mode.

## 2.3 *Synthesis of Stepped-Impedance Transmission Line Structural Elements*

The resonant frequencies $f_1, f_1'$ or $f_1, f_1''$ of dual-mode oscillations in Fig. 1c and 1d form the main passband. Other resonant frequencies are parasitic and should be removed from the main passband. For the resonator in Fig. 1a and 1c it is necessary to increase the ratio of resonant frequencies $f_2/f_1$, and for the resonator in Fig. 1b and 1d it is desirable to increase the ratio $f_3/f_1$. The important element of the considered dual-mode resonators is a horizontal line segment with an electrical length of $2\theta$. It is a half-wave resonator, the resonant frequencies of which are related to each other as natural numbers 1:2:3…. This pattern leads to the value $f_2/f_1 = 2$ for Fig. 1a and to the ratio $f_3/f_1 = 3$ for Fig. 1b.

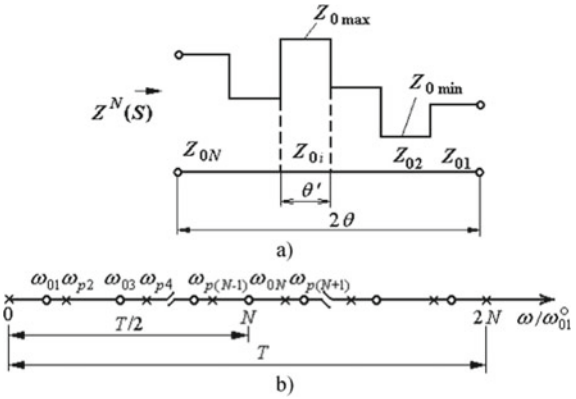



**Fig. 2** SILS with segments of the same length and open-circuited ends: **a** schematic; **b** critical frequencies of input impedance

We will replace this segment by SILS (Fig. 2a) containing $N$ segments of the same length. The electrical length of the entire section is $2\theta$, and the electrical length of one small section is $\theta' = 2\theta/N$. When using the Richard's frequency variable [38]

$$S = j\Omega = j\tan\theta' = j\tan 2\theta/N, \tag{11}$$

the transfer matrix of SILS is written as [39]:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \left(1 - S^2\right)^{-N/2} \times \begin{bmatrix} a_0 + a_2S^2 + \cdots + a_NS^N & a_1S + a_3S^3 + \cdots + a_{N-1}S^{N-1} \\ b_1S + b_3S^3 + \cdots + b_{N-1}S^{N-1} & b_0 + b_2S^2 + \cdots + b_NS^N \end{bmatrix}, \tag{12}$$

where $N$ is an even natural number. If $N$ is an odd number, then the highest degrees of the polynomials are swapped; $a_0 = b_0 = 1$; and the remaining coefficients of the polynomials are positive numbers depending on the characteristic impedances $Z_{0i}$, $i = 1, 2, \ldots, N$. It follows from (12) that the input impedance of the considered SILS vs frequency variable $S$ (11) is described by the reactance function of degree $N$

$$Z^N(S) = \frac{A}{C} = \frac{a_0 + a_2S^2 + \cdots + a_NS^N}{b_1S + b_3S^3 + \cdots + b_{N-1}S^{N-1}}. \tag{13}$$

To obtain the elements of the SILS transfer matrix $[A^{(N)}B^{(N)}C^{(N)}D^{(N)}]$ of order $N$, we need the SILS transfer matrix $[A^{(N-1)}B^{(N-1)}C^{(N-1)}D^{(N-1)}]$ of order $(N - 1)$ multiply on the left by the transfer matrix of the transmission line segment with characteristic impedance $Z_{0N}$.

$$\begin{bmatrix} A^{(N)} & B^{(N)} \\ C^{(N)} & D^{(N)} \end{bmatrix} = \left(1 - S^2\right)^{-1/2} \begin{bmatrix} 1 & SZ_{0N} \\ S/Z_{0N} & 1 \end{bmatrix} \begin{bmatrix} A^{(N-1)} & B^{(N-1)} \\ C^{(N-1)} & D^{(N-1)} \end{bmatrix}. \quad (14)$$

The expression (14) is recurrent. It is used to determine the polynomials coefficients of the transfer matrix (12) with different values of $N$. These coefficients for $N \leq 4$ are presented in Table 3 and Table 4 (Appendix 6.1).

Figure 2b shows the distribution of poles and zeros of the input impedance (13), which are the resonant and anti-resonant frequencies of the SILS. These frequencies are also called critical. The frequency axis is normalized relative to the first antiresonant frequency $\omega^{\circ}_{01}$ of the half-wave uniform resonator. In this case the critical frequencies of this resonator coincide with natural numbers of the series 1:2:3…

These frequencies alternate with each other. They are located periodically with a period $T = 2N$, which corresponds to the electrical length $\theta' = \pi$ of the composite segment. They have an inter-period symmetry. In the case of continuous numbering, the frequency with the number $N$ is the center of symmetry, it can be either zero or pole of $Z^N$, depending on the parity of the number $N$. It does not depend on the characteristic impedances of the composite segments $Z_{0i}$, since it corresponds to the electrical length $\theta' = \pi/2$, at which the frequency variable $S = j\infty$. The first $(N-1)$ critical frequencies completely define the SILS frequency spectrum. The frequencies of the right half of the first period are the mirroring of the frequencies of the left half of the period, relative to the center frequency with the value $\theta' = \pi/2$. Determination of the SILS critical frequencies from the given characteristic impedances $Z_{0i}$ is usually called the analysis problem.

Since the given relationship between two initial critical frequencies (Fig. 2c) can be realized on a set of functions of the characteristic impedance $\{Z_{0k}(x)\}$, $k = 1$, 2,…, then the limited value of the parameter $m = Z_{0\max}/Z_{0\min}$ leads to the following optimization problem. It is required to determine a function $Z_0^*(x)$ among the set of functions $\{Z_{0k}(x)\}$, which provides the resonator with a given ratio between the initial critical frequencies, and at the same time has a minimum value of $m = m_{\min}$. This problem assumes optimization of the resonator shape.

Analysis of the optimal solutions obtained for SILS [46] showed that they are characterized by a twofold degeneration: the minimum (for this problem) number of composite segments $N$ and only two levels of characteristic impedances $Z_{0\min}$ and $Z_{0\max}$. This demonstrates the striving of optimal SILS to take the simplest forms. In [46], a parametric synthesis method was proposed that allows us to minimize $m$ in a simple way. This method is based on the characteristics noted above inherent in optimal SILS, which can be represented as mathematical expressions:

$$Z^N(S) = Z^{N\min}(S), \quad (15)$$

$$Z_{0i} = Z_{0\min} \quad \text{or} \quad Z_{0\max}. \quad (16)$$

Condition (15) means that the order of the realized reactance function $Z^N(S)$ (13) should be minimal. Condition (16) is self-explanatory. In what follows, we will denote $Z_{0\min} = Z_0$ and $Z_{0\max} = mZ_0$.

The parametric synthesis problem can be divided into three stages. At the first step, the minimum order $N_{\min}$ of the reactance function $Z^N(S)$ of the input impedance is determined. In the second step, the function $Z^N(S)$ is expressed in terms of the characteristic impedances $Z_{0i}$ of the SILS (Appendix 6.1). At the third step, the values of $Z_{0i}$ are determined in such a way as to provide the required ratio between two critical frequencies. In this case, the feature (16) is taken into account.

**SILS No. 1.** The frequency response of the dual-mode resonator (Fig. 1c) suggests the use of SILS and distancing from each other the resonance frequencies $\omega_{p1}$ and $\omega_{p2}$. In Fig. 2b, continuous numbering of the SILS critical frequencies is used and the mentioned resonance frequencies are indexed as $\omega_{p2}$ and $\omega_{p4}$, respectively. The main feature when choosing the minimum order $N_{\min}$ of the reactance function of the input impedance $Z^N(S)$ is the inter-period symmetry of its critical frequencies location. Since it is required to distance two lowest resonant frequencies from each other, then $N_{\min} = 3$ should be used. In this case, a decrease in the first of them will lead to an increase in the second one, and their ratio will increase. For $N = 3$, the input impedance function (13) is written as

$$Z^3(S) = \frac{a_0^{(3)} + a_2^{(3)} S^2}{b_1^{(3)} S + b_3^{(3)} S^3} \tag{17'}$$

where $S = j\Omega = j\tan\theta' = j\tan(2\theta/3)$.

We introduce into function (17′) the characteristic impedances (Appendix 6.1)

$$Z^3(S) = \frac{1 + \left(Z_{02}/Z_{01} + Z_{03}/Z_{01} + Z_{03}/Z_{02}\right)S^2}{\left(1/Z_{01} + 1/Z_{02} + 1/Z_{03}\right)S + \left(Z_{02}/Z_{01}Z_{03}\right)S^3}. \tag{17''}$$

The pole of function (17) is located at the frequency

$$\Omega_{p1}^2 = b_1^{(3)} \Big/ b_3^{(3)} = Z_{01}/Z_{02} + Z_{01}Z_{03}/Z_{02}^2 + Z_{03}/Z_{02} \tag{18}$$

Its decrease leads to an increase in the value of $\Omega^2{}_{p2}$ and to an increase in the ratio $\omega_{p2}/\omega_{p1}$. It follows from (18) that the value of $\Omega^2{}_{p1}$ will be minimal if $Z_{01} = Z_{03} = Z_{0\min} = Z_0$ and $Z_{02} = Z_{0\max} = mZ_0$. The synthesized SILS is shown in Fig. 3a. Its input impedance is follows:

$$Z^3(S) = Z_0 \frac{1 + \left(1 + m + 1/m\right)S^2}{\left(2 + 1/m\right)S + mS^3}, \tag{19}$$

and critical frequencies are shown in Fig. 3b.

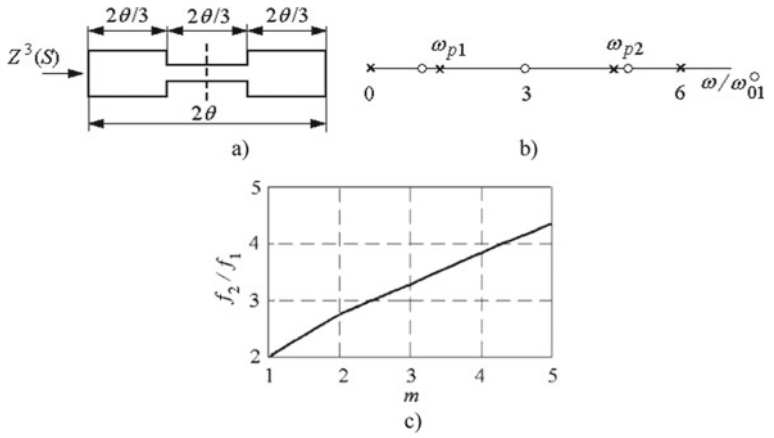



**Fig. 3** SILS No. 1 with increased ratio $f_2/f_1$ of resonance frequencies: **a** Topology, **b** Distribution of critical frequencies, **c** Ratio $f_2/f_1 = f_2/f_1(m)$

From (19) it follows that the ratio $f_2/f_1$ of resonance frequencies depends on the parameter $m$

$$\frac{f_2}{f_1} = \frac{\pi}{\tan^{-1}\sqrt{2/m + 1/m^2}} - 1. \tag{20}$$

The dependence (20) is shown in Fig. 3c.

The parameter

$$\eta = f_1/f_1^{\circ} = 2\theta/\pi. \tag{21}$$

is important. It represents the ratio of the first resonant frequency $f_1$ of SILS to the first resonant frequency of the uniform half-wave segment $f_1^{\circ}$. If $\eta < 1$, then this indicates about shortening of SILS relative to the half-wave resonator when their resonant frequencies coincide.

If $\eta > 1$, then this indicates on increase in the operating frequency of SILS relative to the half-wave resonator when their length coincides.

For the SILS under consideration, from (19) it follows

$$\eta = \frac{3\tan^{-1}\sqrt{2/m + 1/m^2}}{\pi}. \tag{22}$$

If $m = 5$, then from (22) we obtain the shortening degree of the SILS relative to the uniform half-wave segment: $\eta = 0.559$. For $m = 5$, expression (20) gives $f_2/f_1 = 4.36$, and for a uniform segment we have $f_2/f_1 = 2$.

In Fig. 3a, the vertical dashed line divides SILS No. 1 into two equal parts. The transfer matrix of its left half is written as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1+(1+2m)S^2 & Z_0\left[(2+m)S+mS^3\right] \\ Z_0^{-1}\left[\left(2+1/m\right)S+\left(1/m\right)S^3\right] & 1+\left(1+2/m\right)S^2 \end{bmatrix}. \tag{23}$$

where $S = j\tan(\theta/3)$. When composing matrix (23) Appendix 6.1 and (12) were used.

**SILS No. 2.** The frequency response of the dual-mode resonator (Fig. 1d) suggests the use of SILS and distancing from each other the resonance frequencies $\omega_{p1}$ and $\omega_{p3}$. In Fig. 2b, continuous numbering of the critical frequencies of SILS is used, and the mentioned resonance frequencies are indexed as $\omega_{p2}$ and $\omega_{p6}$, respectively. In this case, the minimum order of the reactance function $N_{\min} = 4$ should be used, since a decrease in the first resonance frequency will automatically lead to an increase in the second one and an increase in the ratio of these frequencies. For $N = 4$, the input impedance function (13) is written as

$$Z^4(S) = \frac{a_0^{(4)} + a_2^{(4)}S^2 + a_4^{(4)}S^4}{b_1^{(4)}S + b_3^{(4)}S^3}, \tag{24}$$

where $S = j\Omega = j\tan\theta' = j\tan(\theta/2)$.

The pole of function (24) is located at the frequency.

$$\Omega_{p1}^2 = b_1^{(4)}/b_3^{(4)}, \tag{25}$$

where

$$b_1^{(4)} = 1/Z_{01} + 1/Z_{02} + 1/Z_{03} + 1/Z_{04}, \tag{26}$$

$$\begin{aligned} b_3^{(4)} &= \frac{Z_{02}}{Z_{01}Z_{03}} + \frac{Z_{02}}{Z_{01}Z_{04}} + \frac{Z_{03}}{Z_{01}Z_{04}} + \frac{Z_{03}}{Z_{02}Z_{04}} \\ &= \frac{1}{Z_{01}Z_{04}}\left[\frac{Z_{02}(Z_{03}+Z_{04})}{Z_{03}} + \frac{Z_{03}(Z_{01}+Z_{02})}{Z_{02}}\right]. \end{aligned} \tag{27}$$

A decrease in ${\Omega_{p1}}^2$ leads to an increase in the value of ${\Omega^2}_{p3}$ and to an increase in the ratio $\omega_{p3}/\omega_{p1}$. To decrease $\Omega_{p1}$ (25), it is necessary to decrease the coefficient $b_1$ (26) and increase $b_3$ (27). The maximum effect is achieved at $Z_{01} = Z_{04} = Z_{0\min} = Z_0$ and $Z_{02} = Z_{03} = Z_{0\max} = mZ_0$. The synthesized SILS No. 2 is shown in Fig. 4a. Its input impedance is written as

$$Z^4(S) = Z_0\frac{1+2\left(1+m+1/m\right)S^2+S^4}{2\left(1+1/m\right)S+2(1+m)S^3} \tag{28}$$

and its critical frequencies are shown in Fig. 4b.

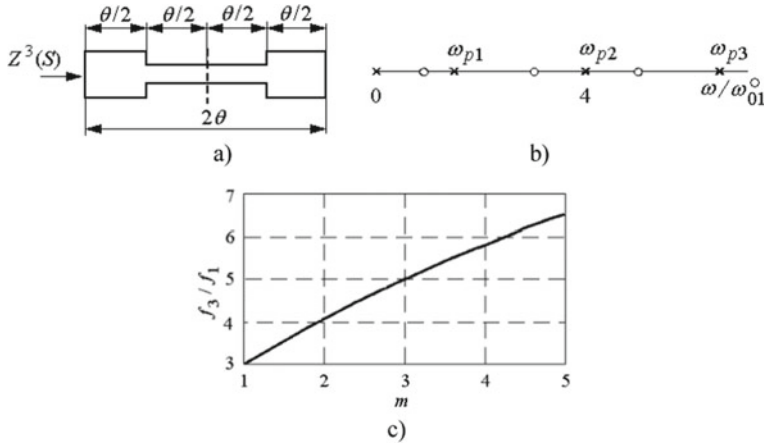



**Fig. 4** SILS No. 2 with increased ratio $f_3/f_1$ of resonance frequencies: **a** Topology, **b** Distribution of resonance frequencies, **c** Ratio $f_3/f_1 = f_3/f_1(m)$

From (28) it follows that the ratio of resonance frequencies $f_3/f_1$ depends on parameter $m$

$$\frac{f_3}{f_1} = \frac{\pi}{\tan^{-1}\sqrt{1/m}} - 1. \tag{29}$$

The relationship (29) is shown in Fig. 4c. It follows from (28) that the parameter $\eta$ (21), which is important for SILS, has the form

$$\eta = \frac{4\tan^{-1}\sqrt{1/m}}{\pi}. \tag{30}$$

If $m = 5$, then from (30) we obtain the shortening degree of the SILS relative to the uniform half-wave segment: $\eta = 0.535$. In the same time, for $m = 5$, (29) gives $f_3/f_1 = 6.47$, while for a uniform half-wave segment we have $f_3/f_1 = 3$.

In Fig. 4a, a vertical dashed line divides SILS No. 2 into two equal parts. The transfer matrix of its left half is written as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 + (1/m)S^2 & Z_0(1+m)S \\ Z_0^{-1}(1+1/m)S & 1 + mS^2 \end{bmatrix} \tag{31}$$

where $S = j\tan(\theta') = j\tan(\theta/2)$.

When compiling matrix (31), Appendix 6.1 and (12) were used.

## 2.4 Dual-Mode Resonators with SILSs

Dual-mode resonator with SILS No. 1. This resonator is shown in Fig. 5a. As before, its resonant frequencies are determined by Eqs. (3). Taking into account the elements of transfer matrix (23), these equations take the form.

$$d = 1 + (1 + 2/m)S^2 = 0 \tag{32'}$$

$$Y_s(S^*) + 2c/d = 0, \tag{32''}$$

where

$$\frac{c}{d} = Z_0^{-1}\frac{(2+1/m)s + (1/m)s^3}{1+(1+2/m)s^2}, \tag{33}$$

and $S = j\tan(\theta/3)$.

The admittance of the stub $Y_s(S^*)$ is expressed by the reactance function $Y_s(S^*) = c/d$ with frequency variable $S^* = j\tan[(\theta \pm \Delta')/3]$. Let's assume that the electrical length of this stub is $\theta' < \theta$. The admittance of the stub $Y^3{}_s(S^*)$ is also expressed by the reactance function (33), and since this stub has a reduced electrical length ($\theta - \Delta'$) compared to the half of SILS No. 1, the frequency variable $S^* = j\tan[(\theta - \Delta')/3]$.

Figure 5b shows the insertion loss frequency responses of the considered dual-mode resonator for weak coupling with I/O loads. A linear simulator of Microwave Office (AWR) was used for plotting. Now and in the future, we will assume $Z_0 = 10$ Ω and $m = 3$.

The resonance Eq. (32′) describes simple oscillations with odd numbers $f_1, f_3, \ldots$ The resonance Eq. (32″) describes simple oscillations with even numbers $f_2, f_4, \ldots$, as well as extreme right-hand oscillations, with odd numbers $f_1$", $f_3$", which forms dual-mode oscillations. The resonator under consideration is remarkable in that it has dual-mode oscillations alternating with single oscillations.

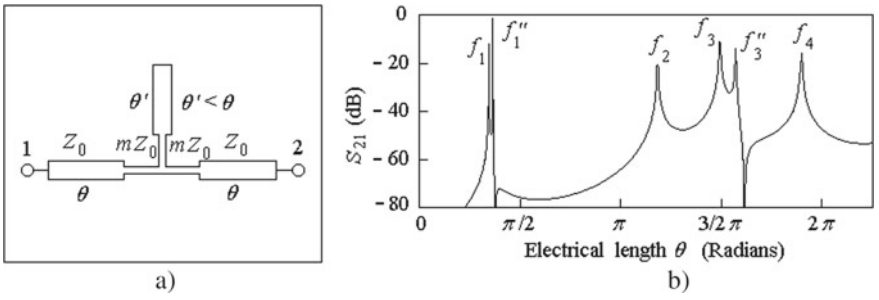



**Fig. 5** Dual mode resonator with SILS No. 1 and $\theta' < \theta$: **a** topology; **b** resonance frequencies

Using SILS No. 1 distances simple oscillation with resonant frequency $f_2$ from the lowest dual-mode oscillation. The distancing of these frequencies from each other is characterized by the ratio $f_2/f_1 = 3.35$, which coincides with the result calculated by formula (20). For $m = 3$, the value $f_1 = 0.69$ GHz, and if $m = 1$, then $f_1{}^\circ = 1$ GHz. Due to the use of SILS No. 1, the main resonance was reduced by 1.45 times, which agrees with the value calculated by (22).

**Dual-mode resonator with SILS No. 2.** This resonator is shown in Fig. 6a, along with SILS No. 2, it contains a short-circuited stub with a small electrical length ($\Delta' << \theta$), which has an input admittance $Y_s = -jZ_0{}^{-1}\cot\Delta'$. As before, the resonant frequencies of the considered resonator (Fig. 6a) are determined by Eqs. (3). Taking into account the elements of transfer matrix (31), these equations take the form:

$$d = 1 + mS^2 = 0 \tag{34'}$$

$$-Z_0^{-1}\cot\Delta' + 2c/d = 0 \tag{34''}$$

where

$$\frac{c}{d} = Z_0^{-1}\frac{(1+1/m)S}{1+mS^2}, \tag{35}$$

and $S = j\tan(\theta/2)$.

Figure 6b shows the insertion loss frequency responses of the dual-mode resonator under consideration with $m = 3$ and weak coupling with I/O loads. The resonance Eq. (34′) describes simple oscillations with odd numbers $f_1, f_3, \ldots$ The resonance Eq. (34″), unlike (32″), does not lead to simple oscillations with even numbers $f_2$, $f_4, \ldots$. It describes only extreme left-handed oscillations with odd numbers, which form dual-mode oscillations. The resonator under consideration is remarkable in that it has only dual-mode oscillations.

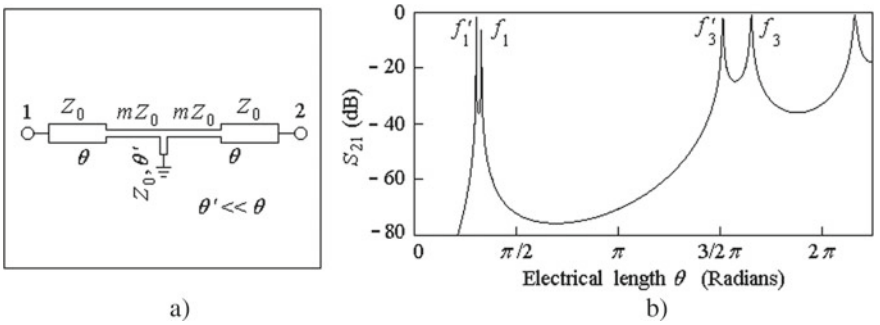



**Fig. 6** Dual-mode resonator with SILS No. 2 and short-circuited stub of short length: **a** topology; **b** resonance frequencies

For $m = 3$, we have $f_1 = 0.667$ GHz and if $m = 1$, then $f_1{}^\circ = 1$ GHz. Due to the use of SILS No. 2, the main resonance was reduced by 1.5 times, which agrees with the value calculated by formula (30). Note that SILS No. 2 is more effective at lowers the main resonance than SILS No. 1.

The use of SILS No. 2 made it possible to distance from each other dual-mode oscillations with central frequencies $f_1$ and $f_3$. The ratio $f_3/f_1$ at $m = 5$ is equal to $f_3/f_1 = 6.47$, which coincides with the value calculated by (29). If in SILS No. 2 the characteristic impedances $Z_0$ and $mZ_0$ are swapped, then in expression (29) it is necessary to replace $m$ by $1/m$. In this case, at $m = 5$, we obtain the ratio $f_3/f_1 = 1.73$, the passband with frequency $f_3$ has significantly approached to the bandwidth with frequency $f_1$. In the filter under consideration, the $f_3/f_1$ ratio can vary over a wide range $1.73 \leq f_3/f_1 \leq 6.47$. This allows this filter to be used as dual-band filter with adjustable position of the second passband relative to the first one.

## 3 Dual-Mode Resonators with Short-Circuited Transmission Line Segments

Consider the symmetric dual-mode resonators shown in Fig. 7a and 7b. The basis of these resonators is a horizontal half-wave transmission line segment short-circuited at both ends, to the middle of which a stub is connected.

### *3.1 Resonance Equation*

A resonance in the symmetrical resonators under consideration is observed when the voltage in the middle of the resonator reaches the maximum or minimum values. Resonance conditions can be expressed in the form:

$$y_{22} = A'/B' = 0 \quad \text{and} \quad 1/y_{22} = B'/A' = 0. \qquad (36'), (36'')$$

Condition (36′) is satisfied when the voltage in the middle of the resonator reaches maximum, which corresponds, in particular, to λ/2-type oscillations.

Condition (36′′) is provided at a minimum voltage in the middle of the resonator. It corresponds, in particular, to λ-type oscillations. If $A'$ and $B'$ do not have coinciding zeros and different poles, then resonance conditions (36) are simplified [49]

$$A' = 0 \quad \text{and} \quad B' = 0. \qquad (37'), (37'')$$

For the considered resonators (Fig. 7), we can use (2), which determines $A'$, $B'$, $C'$, $D'$ parameters of the left half of the resonator. Substituting expressions for $A'$ and $B'$ from (2) in (37), we obtain.

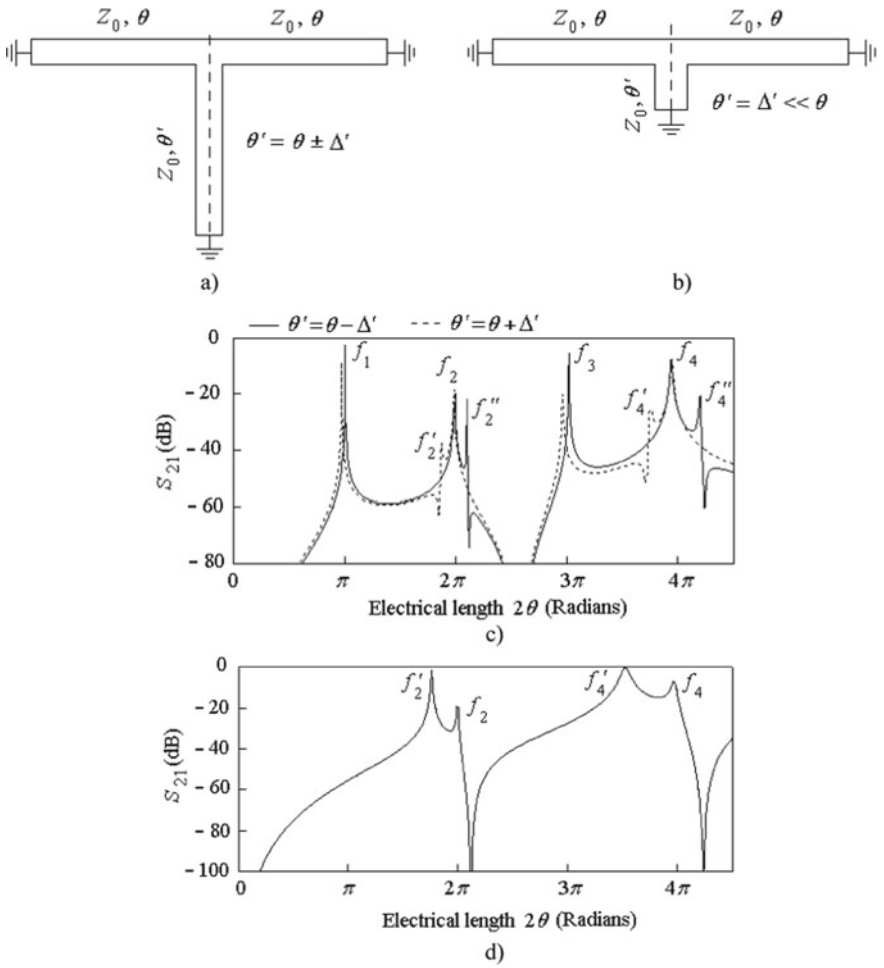



**Fig. 7** Dual-mode resonator with short-circuited ends uniform transmission line segments: **a** topology of resonator with short-circuited stub; **b** topology of resonator with short-circuited stub of short length; **c** resonance frequencies of resonator at Fig. 7a; **d** resonance frequencies of resonator at Fig. 7b

$$b = 0 \quad \text{and} \quad Y_s + 2a/b = 0, \qquad (38'), (38'')$$

where $Y_s$ is the input admittance of the stub at the connection point to the horizontal transmission line segment short-circuited at both ends.

### 3.2 *Properties of Resonators with Uniform Transmission Lines*

In general, the [*abcd*] matrix can represent a segment of non-uniform transmission line. If we use a segment of a uniform transmission line with $Z_0$ and electrical length $2\theta$, then it is necessary to substitute the parameters $a$ and $b$ from matrix (4) into (38):

$$\sin\theta = 0 \quad \text{and} \quad Y_s - j2Z_0^{-1}\cot\theta = 0. \qquad (39'), (39'')$$

Equation (39′) determines the part of the resonant electrical lengths.

$$\theta_n = \pi,\ 2\pi,\ 3\pi, \ldots. \qquad (40)$$

Another part of the resonant electrical lengths is determined by the Eq. (39″) and it depends on the parameters of the used stub. Let us consider the resonator in Fig. 7a in more detail. Its stub has an electrical length $\theta' = \theta \pm \Delta'$, which differs from $\theta$ by a small fixed value $\Delta'$, and the input admittance $Y_s = jZ_0{}^{-1}\ \cot(\theta \pm \Delta')$. The resonance Eq. (39″) for the resonator under consideration takes the form.

$$\cot\left(\theta \pm \Delta'\right) + 2\cot\theta = 0. \qquad (41)$$

Equation (41) has the solutions in the range of electrical lengths $\theta$ for which the function $\cot\theta = 0$ or $\pm\infty$. Figure 7c shows the insertion loss frequency responses of this resonator for weak coupling with I/O loads. It illustrates the resonant frequencies given by Eq. (41). In the first case ($\cot\theta = 0$) the Eq. (41) generates the resonant electrical lengths with odd numbers.

$$\theta_n = \pi/2,\ 3\pi/2,\ 5\pi/2, \ldots \qquad (42)$$

In the second case ($\cot\theta = \pm\infty$), Eq. (41) leads to electrical lengths $\theta_2{}'$, $\theta_2{}''$ and $\theta_4{}'$, $\theta_4{}''$, which are located to the left and to the right from the electrical lengths (40) with even numbers, which gives dual-mode oscillations. If the electrical length of the stub is slightly increased ($\theta + \Delta'$), then extreme left oscillation modes $f_2{}'$, $f_4{}'$, … appear, and with a shorter stub ($\theta - \Delta'$) extreme right oscillation modes $f_2{}''$, $f_4{}''$ appear. The resonator under consideration is remarkable in that it has dual-mode oscillations alternating with single oscillations.

If in Fig. 7a the short-circuited stub with electrical length($\theta \pm \Delta'$) is replaced by a short-circuited stub of short length ($\theta' = \Delta' << \theta$), then we will make the transition to the resonator shown in Fig. 7b. The resonant frequencies of this resonator are shown in Fig. 7d. Part of its resonant electrical lengths is still determined by Eq. (39′) and takes on values (40). In Fig. 7d, they correspond to the resonant frequencies $f_2, f_4$, …. Since the stub admittance $Y_s$ has changed to $Y_s = -jZ_0{}^{-1}\ \cot\Delta'$, the resonance Eq. (39″) takes the form.

$$\cot \Delta' + 2 \cot \theta = 0. \tag{43}$$

The resonance Eq. (43) no longer generates resonant electrical lengths of simple oscillations with odd numbers (42). Equation (43) leads to electrical lengths $\theta_2'$, $\theta_4'$, which are located to the left of the resonant electrical lengths (40) with even numbers, which leads to dual-mode oscillations. The resonator under consideration, and resonator in Fig. 1b, is remarkable in that it contains only dual-mode oscillations.

### *3.3 Synthesis of SILS*

Dual-mode resonators and their frequency responses (Fig. 7) suggest the use of SILS. There are two problems to be solved here. One of them is associated with distancing from each other of simple oscillations with frequencies $f_1$ and $f_3$ (Fig. 7c). In this case, these frequencies will shift away from the main dual-mode oscillation, characterized by the frequencies $f_2', f_2, f_2''$, which will lead to the expansion of the two stopbands located to the right and left of the passband. Another problem involves the distancing from each other of two lower dual-mode oscillations (Fig. 7d), which will lead to an expansion of stopband of the bandpass filter.

To solve the formulated problems, it is necessary to synthesize SILS with a special arrangement of resonant frequencies. The main difficulty is that both ends of the SILS are short-circuited and it does not have an input similar to SILS with two open ends or SILS with one open and one short-circuited ends. The design of SILS is carried out on the input impedance $Z(s)$, which in this case is absent. How to be in this case?

The existing difficulty can be overcome by using the important provisions of the non-uniform transmission line theory [37]. Let a segment of non-uniform transmission line be given with a function of the characteristic impedance $Z_0(x)$ and boundary conditions at the ends corresponding to open circuited, which has a set of resonant frequencies $\{f_{pi}\}$. This set of resonant frequencies $\{f_{pi}\}$ will remain if the previous characteristic impedance $Z_0(x)$ is reversed relative to some constant $Z^*_0(x) = Z^2/Z_0(x)$, and boundary conditions at the ends are replaced by short-circuited ones. This position allows us to carry out parametric synthesis of SILS with two open ends, and in the resulting solution to swap the characteristic impedances $Z_0$ and $mZ_0$.

**SILS No. 3.** For dual-mode resonator (Fig. 7a and 7c) it is required to increase the resonant frequency ratio $f_3/f_1$. This problem has already been solved when using SILS No. 2 with two open ends (Fig. 4a). A line segment with electrical length $2\theta$ is divided into four equal parts, the characteristic impedances of which take the values $Z_{01} = Z_{04} = Z_{0\min} = Z_0$ and $Z_{02} = Z_{03} = Z_{0\max} = mZ_0$. We will use this solution for a line segment that is short-circuited at both ends. Following the pattern [30], let us reverse the indicated characteristic impedances with respect to the value of $mZ_0{}^2$. As a result, we get the desired solution.

$$Z_{01} = Z_{04} = Z_{0\,\max} = mZ_0, \quad Z_{02} = Z_{03} = Z_{0\,\min} = Z_0, \tag{44}$$

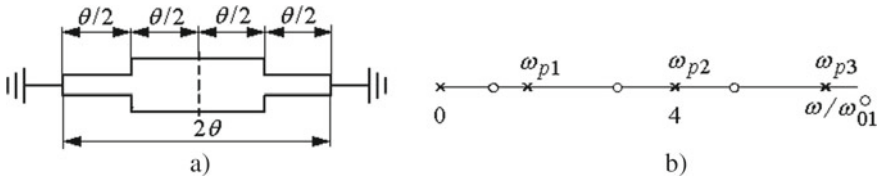



**Fig. 8** SILS No. 3 with increased ratio $f_3/f_1$ of resonance frequencies: **a** topology; **b** distribution of resonance frequencies

which is shown in Fig. 8a and which corresponds to the distribution of resonant frequencies (Fig. 8b).

The dependence of the ratio of resonant frequencies $f_3/f_1$ on parameter $m$ of the synthesized SILS No. 3 (44) with two short-circuited ends, and SILS No. 2 with two open ends, is determined by (29). This dependence is shown in Fig. 4c.

In Fig. 8a, the vertical dashed line divides SILS No. 3 into two equal parts. The transfer matrix necessary for the subsequent presentation of its left half has the form

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 + mS^2 & Z_0(1+m)S \\ Z_0^{-1}\left(1 + 1/m\right)S & 1 + \left(1/m\right)S^2 \end{bmatrix}, \tag{45}$$

where $S = j\tan(\theta') = j\tan(\theta/2)$. When compiling matrix (45), Appendix 6.1 and (12) were used.

**SILS No. 4.** For dual-mode resonator (Fig. 7b and 7d) it is required to increase the resonant frequency ratio $f_4/f_2$. To solve it, at the first step, we use SILS with two open ends (Fig. 2). In Fig. 2b, continuous numbering of the critical frequencies of SILS is used, and the mentioned resonance frequencies are indexed as $\omega_{p4}$ and $\omega_{p8}$, respectively. To distance the indicated frequencies from each other, the reactance function $Z^N(S)$ should be used with a minimum order of $N_{\min} = 3$. In this case, the inter-period symmetry and location periodicity of the critical frequencies of the function $Z^3(S)$ are manifested in the following regularity. To increase the ratio $\omega_{p8}/\omega_{p4}$, it is necessary to lower the value of the resonant frequency $\omega_{p2}$. This will lead to a decrease in the value of $\omega_{p4}$ and to an increase in the value of $\omega_{p8}$, which will be accompanied by an increase in the ratio $\omega_{p8}/\omega_{p4}$. The larger $\omega_{p2}$, the greater the ratio $\omega_{p8}/\omega_{p4}$.

At $N = 3$, the input impedance function $Z_3(S)$ is written as (17′). The pole of this function is located at the frequency (18).

$$\Omega_{p1}^2 = b_1^{(3)}/b_3^{(3)} = Z_{01}/Z_{02} + Z_{01}Z_{03}/Z_{02}^2 + Z_{03}/Z_{02}.$$

According to this expression, the value of $\Omega^2{}_{p1}$ will be maximum if.

$$Z_{01} = Z_{03} = Z_{0\max} = mZ_0, \quad Z_{02} = Z_{0\min} = Z_0. \tag{46}$$

For SILS with two open ends and characteristic impedances (46), the input impedance is written as

$$Z^3(S) = Z_0 \frac{1 + \left(1 + m + 1/m\right)S^2}{\left(1 + 2/m\right)S + \left(1/m^2\right)S^3}. \tag{47}$$

From (47) follows the ratio of resonance frequencies $f_4/f_2$ as the function of parameter $m$

$$\frac{f_4}{f_2} = \frac{\pi + \tan^{-1}\sqrt{m(m+2)}}{\pi - \tan^{-1}\sqrt{m(m+2)}}. \tag{48}$$

Let us reverse the indicated characteristic impedances (46) of SILS with two open ends with respect to the value of $mZ_0{}^2$. As a result, we obtain the desired solution for SILS with two short-circuited ends.

$$Z_{01} = Z_{03} = Z_{0\min} = Z_0, \quad Z_{02} = Z_{0\max} = mZ_0. \tag{49}$$

which is shown in Fig. 9a and corresponds to the distribution of resonant frequencies (Fig. 9b).

The resonance frequency ratio $f_4/f_2$ of SILS No. 4 (49) is determined by (48), which at $m = 5$ gives the value $f_4/f_2 = 2.61$. SILS No. 4 has the same shape as SILS No. 1 (Fig. 3a), although they provide different effects. SILS No. 1 increases the $f_2/f_1$ resonant frequency ratio and SILS No. 4 increases the $f_4/f_2$ resonant frequency ratio. The coincidence of the shapes of these two SILSs allows us to describe the

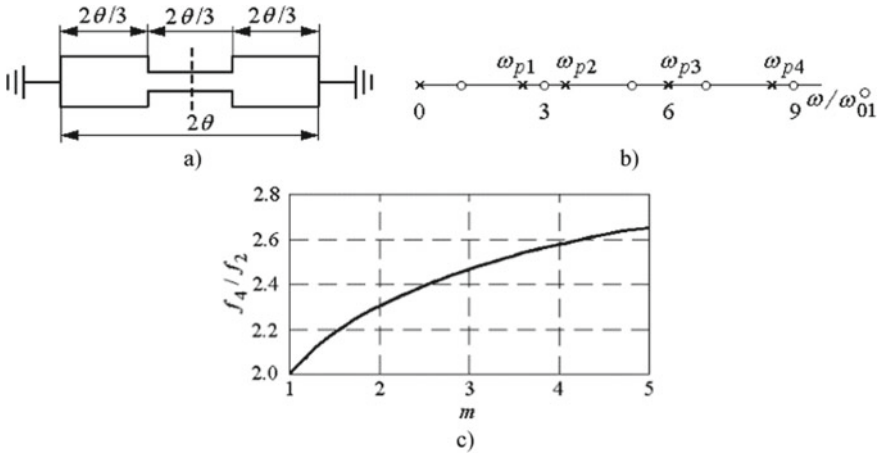



**Fig. 9** SILS No. 4 with increased ratio $f_4/f_2$ of resonance frequencies: **a** topology; **b** distribution of resonance frequencies; **c** ratio $f_4/f_2 = f_4/f_2(m)$

transfer matrix of the left half of SILS No. 4 (Fig. 9a) using the previously obtained expression (23).

### 3.4 Dual-Mode Resonators with SILSs

We will consider two dual-mode resonators that contain different SILSs (Figs. 8 and 9).

**Dual-mode resonator with SILS No. 3.** This resonator is shown in Fig. 10a. As before, its resonant frequencies are determined by Eqs. (38). Taking into account the elements of transfer matrix (45), these equations take the form.

$$b = Z_0(1+m)S = 0, \quad Y_s\left(S^*\right) + 2a/b = 0, \qquad (50'), (50'')$$

where

$$a/b = Z_0^{-1}\left(1 + mS^2\right)/(1+m)S \qquad (51)$$

and $S = j\tan(\theta/2)$. The admittance of the stub $Y_s(S^*)$ is expressed by the reactance function $Y_s(S^*) = a/b$ (51) with frequency variable $S^* = j\tan[(\theta \pm \Delta')/2]$.

This admittance is the reactance function of second degree. The input admittance of the stub (Fig. 10a) at the connection point to SILS No. 3 is designated as $Y^2{}_s(S^*)$. Let's assume that the electrical length of this stub is $\theta' = \theta - \Delta' < \theta$. The admittance $Y^2{}_s(S^*)$ is also expressed by the reactance function (33), and since this stub has a reduced electrical length $(\theta - \Delta')$, the frequency variable $S^* = j\tan[(\theta - \Delta')/2]$ is used for it.

Figure 10b shows the insertion loss frequency responses of the considered dual-mode resonator for reduced electrical length $\theta' = \theta - \Delta'$ of the stub. When plotting, we assumed $Z_0 = 10\ \Omega$ and $m = 3$.

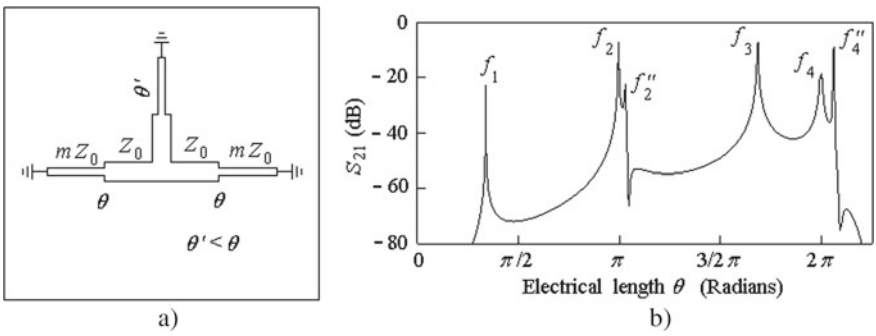



**Fig. 10** Dual-mode resonator with SILS No. 3 and $\theta' < \theta$: **a** topology; **b** resonance frequencies

Equation (50') determines the part of the resonant electrical lengths with even numbers.

$$\theta_n = \pi,\ 2\pi,\ 3\pi, \ldots. \tag{52}$$

The resonant electrical lengths (52) and corresponding resonant frequencies $f_2$, $f_4$, …, do not depend on the parameter $m$ of SILS No. 3.

The resonance Eq. (50″) describes simple oscillations with odd numbers $f_1, f_3$, …, and extreme right-hand oscillations $f_2$", $f_4$", adjacent to simple oscillations (52) with even numbers, which forms dual-mode oscillations. If the stub has an increased length, then resonant frequencies $f_2'$, $f_4'$, … belonging to it, will be located to the left of the oscillations (52) with even numbers.

Using SILS No. 3 allows us to distance simple oscillation with resonant frequencies $f_1$ and $f_3$ from dual-mode oscillation with frequency $f_2$. The distancing of these frequencies from each other is characterized by the ratio $f_3/f_1$, which is determined by (29). For $m = 5$ we have the value $f_3/f_1 = 6.47$, and if $m = 1$, then $f_3/f_1 = 3$.

**Dual-mode resonator with SILS No. 4.** This resonator is shown in Fig. 11a, along with SILS No. 2, it contains a short-circuited stub with a small electrical length $\Delta' << \theta$, which has an input admittance $Y_s = -j(mZ_0)^{-1}\cot\Delta'$. As before, its resonant frequencies are determined by Eqs. (21). Taking into account the elements of its transfer matrix (23), these equations take the form.

$$b = Z_0\left[(2 + m)S + mS^3\right] = 0 \tag{53'}$$

$$-(mZ_0)^{-1}\cot\Delta' + 2a/b = 0, \tag{53''}$$

where.

$$a/b = Z_0^{-1}\left[1 + (1 + 2m)S^2\right]/\left[(2 + m)S + mS^3\right] \tag{54}$$

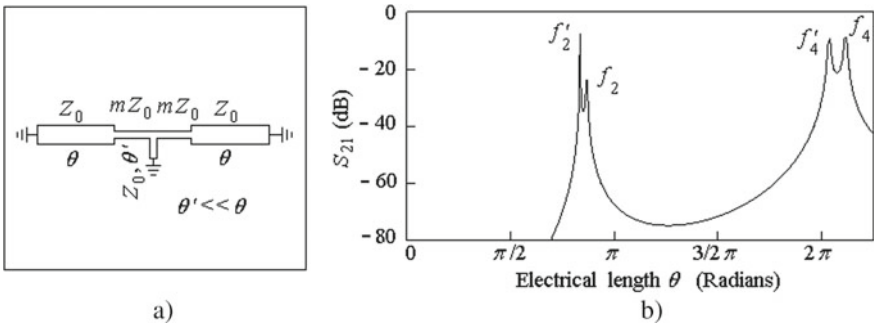



**Fig. 11** Dual-mode resonator with SILS No. 4 and short-circuited stub of short length: **a** topology; **b** resonance frequencies

and $S = j\tan(\theta/3)$.

Figure 11b shows the insertion loss frequency responses of the dual-mode resonator under consideration with parameter $m = 3$ and weak coupling with I/O loads. Equation (53′) determines the part of the resonant electrical lengths with even numbers, which correspond to the resonant frequencies $f_2, f_4, \ldots$. These frequencies depend on parameter $m$ of SILS No. 4 as indicated by Eq. (53′).

The resonance Eq. (53″), unlike the Eq. (50″), does not lead to the simple oscillations with odd numbers $f_1, f_3, \ldots$. It generates only extreme left-hand oscillations $f_2'', f_4'', \ldots$, adjacent to oscillations with even numbers, defined by Eq. (53″), which leads to dual-mode oscillations.

This resonator has only dual-mode oscillations. The resonant electrical length of its lowest dual-mode oscillation is $\theta_2 = \left(3\big/2\right)[\pi - \tan^{-1}\sqrt{m(m+2)}]$.

The parameter $\eta$ (21) in this case is written as.

$$\eta = 3\,[\pi - \tan^{-1}\sqrt{m(m+2)}]/\pi. \tag{55}$$

For $m = 5$, from (55) we obtain the value $\eta = 1.66$. This value indicates that the lowest dual-mode oscillation of the resonator (Fig. 11) is 1.66 times higher in frequency than main oscillation of half-wave resonator.

The use of SILS No. 4 made it possible to distance from each other dual-mode oscillations with central frequencies $f_2$ and $f_4$. For $m = 5$, the $f_4/f_2$ ratio was $f_4/f_2 = 2.61$, which coincides with the value calculated by (48). If in SILS No. 4 the characteristic impedances $Z_0$ and $mZ_0$ are interchanged, then in expression (48) it is necessary to replace $m$ by $1/m$. In this case, at $m = 5$, we obtain the ratio $f_4/f_2 = 1.46$ and value $\eta = 2.44$. In the filter under consideration, the $f_4/f_2$ ratio can vary over a wide range $1.46 \le f_4/f_2 \le 2.61$, which allows this filter to be used as dual-band filter with adjustable position of two passbands.

Note that resonant frequencies of dual-mode resonators with short-circuited ends cannot be extracted using the odd/even analysis (Appendix 6.2).

## 4 Microstrip Dual-Mode Filters with SILSs

To confirm the established patterns, we consider four dual-mode filters with SILSs shown in Figs. 12, 13, 14, 15. The filters are implemented on a substrate of TMM-10i (Rogers), $\varepsilon_r = 9.8$, $\tan\delta = 0.002$, $h = 1.905$ mm. All filters contain structural elements with a maximum width of 3.2 mm ($Z_{0\min} = 37.1\ \Omega$) and minimum width of 0.8 mm ($Z_{0\max} = 70.8\ \Omega$), their parameter $m = 1.91$. The coupling coefficient $K$ between the oscillations of the filters is determined by using EM simulator of the Microwave Office software and the formula (10). The numbering of resonant frequencies (bandwidths) used below will correspond to those used previously in Sects. 2 and 3.

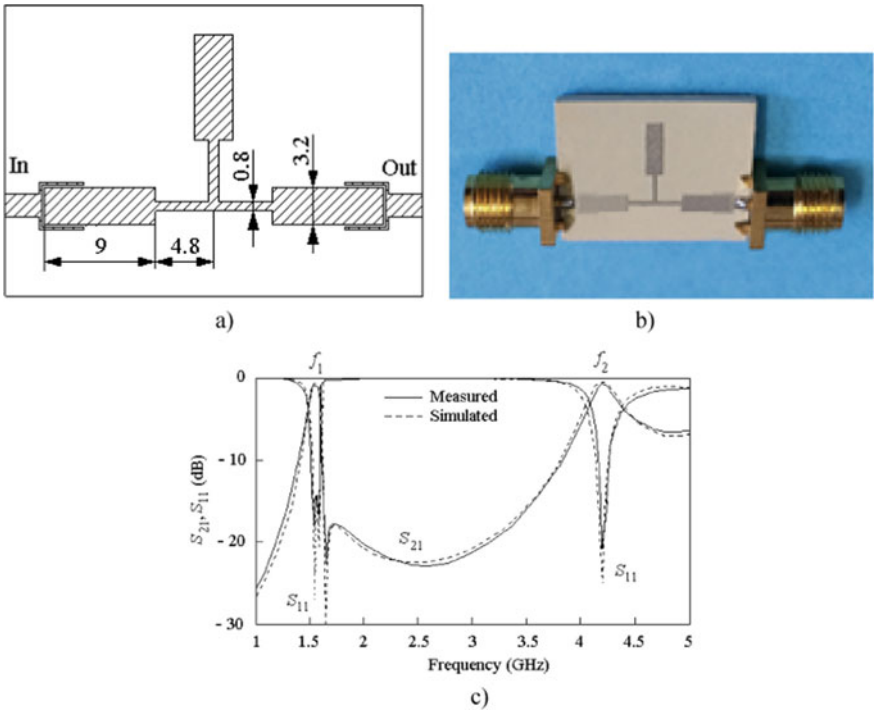



**Fig. 12** Microstrip dual-mode resonator with SILS No: **a** topology; **b** photograph of fabricated filter; **c** measured and simulated frequency responses

## *4.1 Filter with SILS No. 1*

The topology of the filter is shown in Fig. 12a, where dimensions are in millimeters, and it has size 34 mm × 25 mm. The photograph of this filter is shown in Fig. 12b, and its measured and simulated frequency responses are depicted in Fig. 12c. With the specified parameters, the coupling coefficient $K = 0.065$ is provided between oscillations.

The measured parameters are as follows: the center frequency $f_1 = 1575$ MHz; bandwidth $BW = 80$ MHz ($FBW = 0.051$); the minimum insertion loss $IL_{\min} = 0.7$ dB; the return loss in bandwidth $RL > 14$ dB. We use in the filter a shorter stub with electrical length $\theta'$ ($\theta' < \theta$), which resulted in a right-hand transmission zero $f_z = 1647$ MHz. If a longer stub ($\theta' > \theta$) is used, then the transmission zero $f_z$ will be located to the left of the bandpass.

The nearest parasitic bandpass is due to the simple oscillation with frequency $f_2 = 4.2$ GHz. The measurements showed that dual-mode and simple oscillations simultaneously exist in the filter under consideration. The filter has an extended stopband due to an increased $f_2/f_1 = 2.67$ ratio. Calculation by (20) at $m = 1.91$

gives the same value. The filter with uniform transmission line segments has $f_2/f_1 = 2$.

The horizontal part of the resonator, representing SILS No. 1, has length $L = 27.6$ mm. If a uniform transmission line with a width of 3.2 mm ($Z_0 = 37.1\ \Omega$) is used, then the horizontal segment will be half-wave and have a length of $\lambda_0/2 = 36$ mm. With the used value $m = 1.91$, the length of SILS No. 1 turned out to be shorter by $\eta = 0.77$. This shortening was contributed by the capacitances at the input and output of the filter. For $m = 1.91$, (22) gives the value $\eta = 0.81$.

## 4.2 Filter with SILS No. 2

The topology of the filter is shown in Fig. 13a. It occupies a 34 mm × 15 mm substrate. A metallized hole with diameter of 0.6 mm is made in its central part. Simulated frequency responses of the filter are depicted in Fig. 13b. With the specified parameters, the coupling coefficient $K = 0.089$ is provided between oscillations.

The simulated parameters are as follows: the center frequency $f_1 = 1440$ MHz; BW = 115 MHz (FBW = 0.08); the insertion loss at center frequency $IL_0 = 0.5$ dB; the return loss in bandwidth $RL > 14$ dB, selectivity 20 dB ($f_0 \pm 300$ MHz).

This filter has only dual-mode oscillation, which correspond to the odd-numbered passbands (Fig. 13b), $f_3 = 6.12$ GHz. The filter has an extended stopband due to the increased ratio $f_3/f_1 = 4.25$. The filter with uniform transmission line segments has $f_3/f_1 = 3$. In Fig. 13b, the arrows indicate the cutoff frequencies of the stopband at 20 dB level. The rejection band is wide enough and it is characterized by a boundary frequency ratio of 3.02.

The resonator of this filter has a length $L = 27.6$ mm. With the used value $m = 1.91$, the length of SILS No. 2 turned out to be shorter by $\eta = 0.8$ (30).

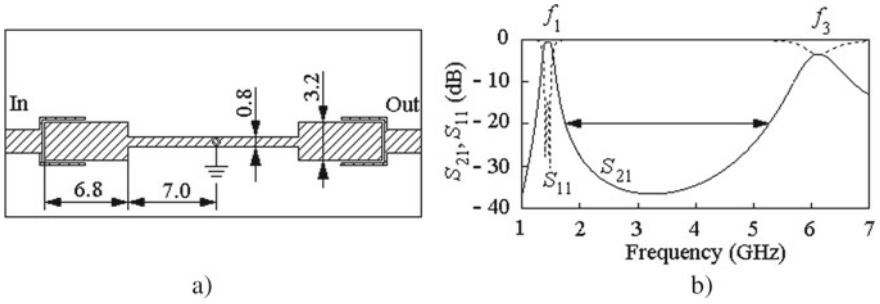



**Fig. 13** Microstrip dual-mode resonators with SILS No: **a** topology; **b** simulated frequency responses

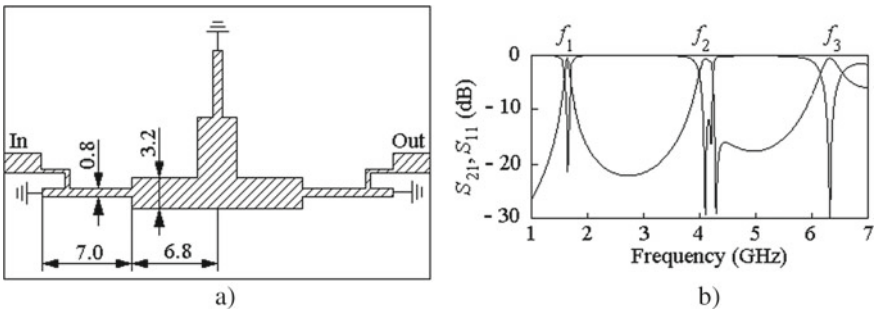



**Fig. 14** Microstrip dual-mode resonators with SILS No. 2: **a** topology; **b** simulated frequency responses

### *4.3 Filter with SILS No. 3*

The topology of the filter is shown in Fig. 14a, and it has size 34 mm × 23 mm. Its three ends are short-circuited with metalized holes. The simulated frequency responses of the filter are depicted in Fig. 14b. With the specified parameters, the coupling coefficient $K = 0.055$ is provided between oscillations.

Its three passbands are located at $f_1 = 1.639$ GHz, $f_2 = 4.12$ GHz and $f_3 = 6.3$ GHz. The passbands at frequencies $f_1$ and $f_3$ are spurious and correspond to single oscillations. A useful bandpass is $f_2$, which is formed by dual-mode oscillations.

Its simulated parameters are as follows: the center frequency $f_2 = 4.12$ GHz; BW = 180 MHz (FBW = 0.044); the minimum insertion loss $IL_{\min} = 0.6$ dB; the return loss in bandwidth $RL > 13$ dB. We use in the filter a shorter stub with electrical length $\theta'$ ($\theta' < \theta$), which resulted in a right-hand transmission zero $f_z = 4.3$ GHz. If a longer stub ($\theta' > \theta$) is used, then the transmission zero $f_z$ will be located to the left of the bandpass.

The use of SILS No. 3 made it possible to increase the $f_3/f_1$ ratio from $f_3/f_1 = 3$ to $f_3/f_1 = 3.84$, which slightly widened both stopbands located to the left and right of $f_2$. The presence of a parasitic passband at frequency $f_1$ is a significant drawback and does not make this filter attractive for practical use. This filter clearly illustrates the alternation of dual-mode and simple oscillations in a distributed circuit.

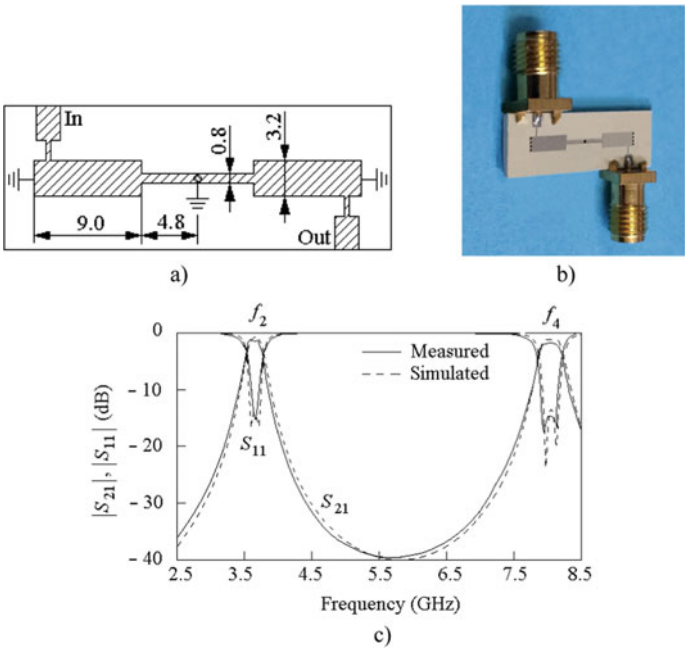



**Fig. 15** Microstrip dual-mode resonators with SILS No. 4: **a** topology; **b** photograph of fabricated filter; **c** measured and simulated frequency responses

## *4.4 Filter with SILS No. 4*

The topology of the filter is shown in Fig. 15a. It occupies a 31.6 mm × 12.8 mm substrate. In its central part, a metallized hole with a diameter of 0.6 mm is made. Its two ends are short-circuited using metalized holes with a diameter of 0.4 mm. The photograph of this filter is shown in Fig. 15b, and its measured and simulated frequency responses are depicted in Fig. 15c. With the specified parameters, the coupling coefficient $K = 0.048$ is provided between oscillations. Similar to the filter in Fig. 13, this filter can be used as a dual-band filter. Table 1 compares the experimental data of this filter with other refereed dual-band BPFs. The measured ratio of the center frequencies of the two passbands was $f_2/f_1 = 2.21$. The calculated ratio of the same frequencies using (48) at $m = 1.91$ is equal to $f_2/f_1 = 2.26$. The difference between these values is 2%. This dual-band filter is notable for the fact that by changing the parameter m we can significantly change the relative position of two passbands $1.46 \leq f_2/f_1 \leq 2.61$. Another positive feature of this filter is that it operates at higher operating frequencies. The central frequency $f_1 = 3.64$ GHz of its lower passband at $m = 1.91$ is increased by $\eta = 1.83$ times (55).

Let's systematize the properties of dual-mode resonators with synthesized SILSs. Based on the optimization method [37], two pairs of SILSs were synthesized, which led to four new dual-mode resonators, the properties of which are systematized in

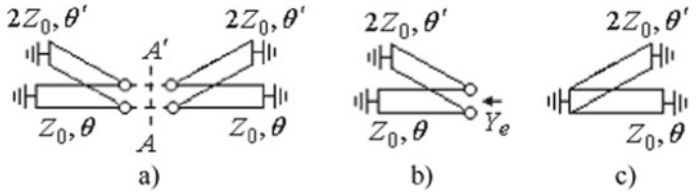



**Fig. 16** Dual-mode resonator with short-circuited ends: **a** symmetrical network presentation; **b** even-mode excitation; **c** odd-mode excitation

**Table 1** Comparison with referred dual-mode filters

| Refs | First band | | | Second band | | | $f_2/f_1$ |
|---|---|---|---|---|---|---|---|
| | $f_1$ (GHz) | 3-dB FBW | $IL_0$ (dB) | $f_2$ (GHz) | 3-dB FBW | $IL_0$ (dB) | |
| [15] | 0.705 | 0.41 | 0.55 | 1.37 | 0.19 | 1.04 | 1 |
| [16] | 3.61 | 0.21 | 0.65 | 6.14 | 0.17 | 1.34 | 1 |
| [17] | 1.78 | 0.07 | 0.42 | 4.0 | 0.055 | 0.37 | 1 |
| This work | 3.64 | 0.07<br>255 | 0.8 | 8.06 | 0.045<br>365 | 1.2 | 2.21<br>$1.46 \leq f_2/f_1 \leq 2.61$ |

**Table 2** Properties of dual-mode resonators with SILS

| SILSs | Single Oscillations | Dual-band BPF | Ratio $f_2/f_1$ ($f_3/f_1$) | $\eta$ |
|---|---|---|---|---|
| SILS No.1 | Yes | No | $2 < f_2/f_1 \leq 4.36$ | $0.558 \leq \eta \leq 1$ |
| SILS No.2 | No | Yes | $1.73 \leq f_2/f_1 \leq 6.47$ | $0.535 \leq \eta \leq 1$ |
| SILS No.3 | Yes | No | $3 \leq f_3/f_1 \leq 6.47$ | - |
| SILS No.4 | No | Yes | $1.46 \leq f_2/f_1 \leq 2.61$ | $1.66 \leq \eta \leq 2.44$ |

the summary Table 2. The continuous numbering of resonant frequencies is used in this table.

The dual-mode resonator with SILS No. 1, has an extended stopband, since the $f_2/f_1$ ratio of the center frequencies of the parasitic and main passbands can reach the value $f_2/f_1 = 4.36$. This resonator is much shorter than the half-wave one. Its shortening factor $\eta$ can be the value $\eta = 0.558$.

The dual-mode resonator with SILS No. 2, has a wide rejection band, since the $f_2/f_1$ ratio can reach $f_2/f_1 = 6.47$. This resonator has the maximum shortening $\eta = 0.535$. This resonator is very promising for use in dual-band BPF, since it allows to vary within wide limits $1.73 \leq f_2/f_1 \leq 6.47$ the relative position of two passbands.

The dual-mode resonator with SILS No. 3 is characterized by the fact that its simple oscillations with frequencies $f_1$, $f_3$,…, and dual-mode oscillations with frequencies $f_2, f_4$,…, alternate with each other. The SILS No. 3 allows us to distance parasitic simple oscillations with frequencies $f_1$ and $f_3$ from the main dual-mode

oscillation with frequency $f_2$, and expand the stopbands located to the right and left of $f_2$. The ratio $f_3/f_1$ can reach the value $f_3/f_1 = 6.47$.

The dual-mode resonator with SILS No. 4 has an extended stopband. Its $f_2/f_1$ ratio can reach the value $f_2/f_1 = 2.61$. This resonator is notable for the fact that its main resonant frequency is located significantly higher than the resonant frequency of half-wave resonator ($1.66 \leq \eta \leq 2.44$), which makes it possible to increase the operating frequencies. This resonator is promising for use in dual-band BPF. The relative position of its passbands can vary within a wide range of $1.46 \leq f_2/f_1 \leq 2.61$.

Note that when forming a conductive BPF pattern by etching under the influence of a magnetic field, it is necessary to take into account the effects presented in [54]. The considered filters can find application in the equipment of the millimeter wave range [55].

## 5 Conclusion

SILS No.1- SILS No.4 synthesized by parametric synthesis allowed expanding the functionality of dual-mode BPFs, some of which (SILS No. 2 and SILS No. 4) can be used as dual-band filters with a wide variable $f_2/f_1$ ratio. The synthesized SILSs are the best among all kinds of smoothly-impedance and stepped-impedance transmission line segments. Consideration of dual-mode resonators with two short-circuited ends (SILS No. 3 and SILS No. 4) was previously difficult for two reasons related to circuit theory. First, the traditional odd / even analysis is not applicable to them. Secondly, there was no method for synthesizing stepped-impedance transmission line segments with two short-circuited ends at specified resonant frequencies. These difficulties are eliminated in this article. Also new in circuit theory is the property of alternating dual-mode and simple oscillations in some resonators (SILS No. 1 and SILS No. 3). The results obtained can be extended to triple-mode resonators. Some of them with characteristics are presented in Figs. 17, 18, 19, 20 (Appendix 6.3).

# 6 Appendix

## 6.1 Transfer Matrix Polynomial Coefficients of SILS

**Table 3** Transfer matrix polynomial coefficients of SILS

| $N$ | $a_0^{(N)}, b_0^{(N)}$ | $a_1^{(N)}, b_1^{(N)}$ | $a_2^{(N)}, b_2^{(N)}$ |
|---|---|---|---|
| 1 | $a_0^{(1)} = 1\ b_0^{(1)} = 1$ | $a_1^{(1)} = Z_{01}$<br>$b_1^{(1)} = 1/Z_{01}$ | |
| 2 | $a_0^{(2)} = 1\ b_0^{(2)} = 1$ | $a_1^{(2)} = Z_{01} + Z_{02}$<br>$b_1^{(2)} = 1/Z_{01} + 1/Z_{02}$ | $a_2^{(2)} = Z_{02}/Z_{01}$<br>$b_2^{(2)} = Z_{01}/Z_{02}$ |
| 3 | $a_0^{(3)} = 1\ b_0^{(3)} = 1$ | $a_1^{(3)} = Z_{01} + Z_{02} + Z_{03}$<br>$b_1^{(3)} = 1/Z_{01} + 1/Z_{02} + 1/Z_{03}$ | $a_2^{(3)} = Z_{02}/Z_{01} + Z_{03}/Z_{01} + Z_{03}/Z_{02}$<br>$b_2^{(3)} = Z_{01}/Z_{02} + Z_{01}/Z_{03} + Z_{02}/Z_{03}$ |
| 4 | $a_0^{(4)} = 1$<br>$b_0^{(4)} = 1$ | $a_1^{(4)} = Z_{01} + Z_{02} + Z_{03} + Z_{04}$<br>$b_1^{(4)} = 1/Z_{01} + 1/Z_{02} + 1/Z_{03} + 1/Z_{04}$ | $a_2^{(4)} = Z_{02}/Z_{01} + Z_{03}/Z_{01} + Z_{04}/Z_{01} + Z_{03}/Z_{02} + Z_{04}/Z_{02} + Z_{04}/Z_{03}$<br>$b_2^{(4)} = Z_{01}/Z_{02} + Z_{01}/Z_{03} + Z_{01}/Z_{04} + Z_{02}/Z_{03} + Z_{02}/Z_{04} + Z_{03}/Z_{04}$ |

## 6.2 Dual-Mode Resonator with Short-Circuited Ends

Dual-mode resonator with short-circuited ends (Fig. 7a and 7b) can be represented as a symmetrical network in Fig. 16a, where AA′ is a symmetrical plane of the network. If the open circuit is AA′ plane, then we get a circuit (Fig. 16b) for even mode analysis. In this case, the resonance equation $Y_e = 0$ can be drawn up, which will give the resonance frequencies of even oscillation modes. If the short circuit is AA′ plane, then we will get a circuit (Fig. 16c) for odd mode oscillations. The resonator in Fig. 16c has no input, so the resonance equation for odd mode oscillations cannot be drawn up (Tables 3 and 4).

## 6.3 Triple Mode Resonator with SILS

**Table 4** Continuation transfer Matrix Polynomial Coefficients of SILS

| $N$ | $a_3^{(N)}, b_3^{(N}$ | $a_4^{(N)}, b_4^{(N}$ |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | $a_3^{(3)} = Z_{01}Z_{03}/Z_{02}$<br>$b_3^{(3)} = Z_{02}/Z_{01}Z_{03}$ | |
| 4 | $a_3^{(4)} = Z_{01}Z_{03}/Z_{02} + Z_{01}Z_{04}/Z_{02} + Z_{01}Z_{04}/Z_{03} + Z_{02}Z_{04}/Z_{03}$<br>$b_3^{(4)} = Z_{02}/Z_{01}Z_{03} + Z_{02}/Z_{01}Z_{04} + Z_{03}/Z_{01}Z_{04} + Z_{03}/Z_{02}Z_{04}$ | $a_4^{(4)} = Z_{02}Z_{04}/Z_{01}Z_{03}$<br>$b_4^{(4)} = Z_{01}Z_{03}/Z_{02}Z_{04}$ |

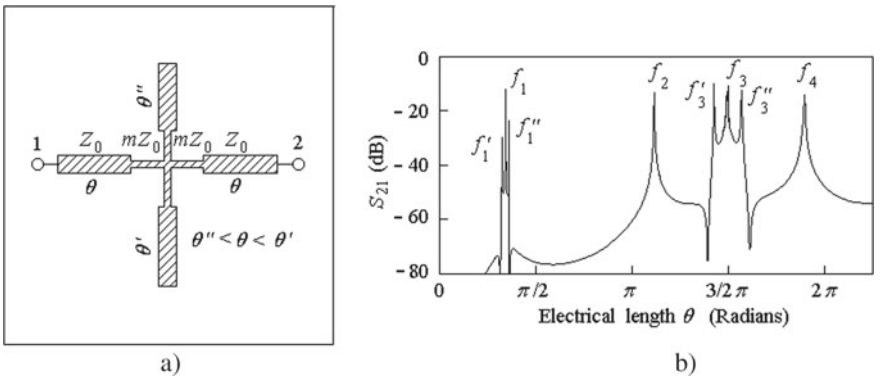



**Fig. 17** Triple mode resonator with SILS No. 1: **a** topology; **b** resonance frequencies (increased ratio $f_2/f_1$, reduced $f_1$)

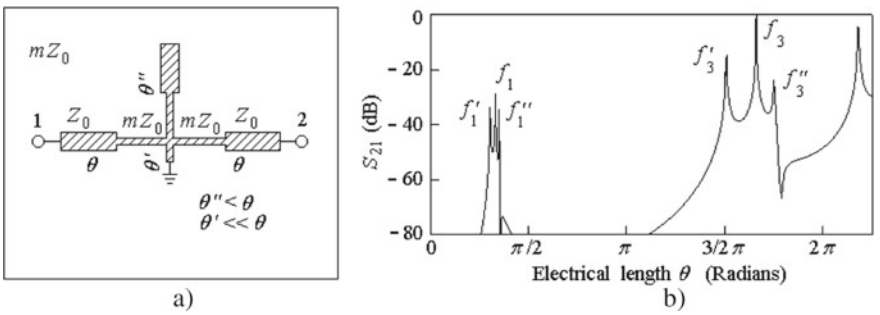



**Fig. 18** Triple mode resonators with SILS No. 2 and short-circuited stub of short length: **a** topology; **b** resonance frequencies (increased ratio $f_3/f_1$, reduced $f_1$)

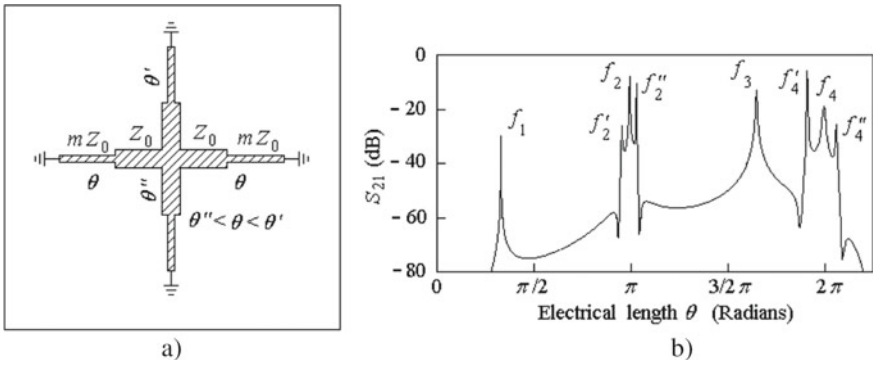



**Fig. 19** Triple mode resonator with SILS No. 3: **a** topology; **b** resonance frequencies (increased ratio $f_3/f_1$)

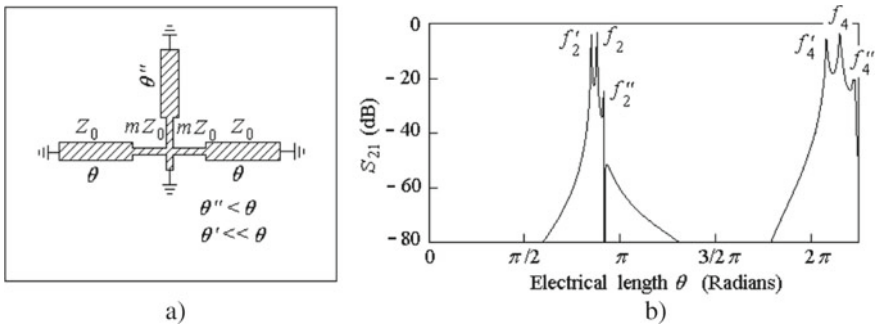



**Fig. 20** Triple mode resonators with SILS No. 4 and short-circuited stub of short length: **a** topology; **b** resonance frequencies (increased ratio $f_4/f_2$, increased $f_2$)

## References


1. Zakharov A (2021) Transmission zeros of trisection and quadruplet bandpass filters with mixed cross coupling. IEEE Trans Microw Theory Techn 1(69):89–100
2. Perodou A, Korniienko A, Scorletti G, Zarudniev M, David J-B, O'Connor I (2021) Frequency design of lossless passive electronic filters: a state-space formulation of the direct synthesis approach. IEEE Trans Circuits Syst I Reg Papers 1(68):161–174
3. Zakharov A, Ilchenko M (2020) Coupling coefficients between resonators in stripline combline and pseudocombline bandpass filters. IEEE Trans Microw Theory Techn 7(68):2679–2690
4. Zakharov A, Ilchenko M (2020) Circuit function characterizing tunability of resonators. IEEE Trans Circuits Syst I Reg Papers 1(67):98–107
5. Zakharov A, Ilchenko M (2020) Unloaded quality factor of transmission line resonators with capacitors. IEEE Trans Circuits Syst I Reg Papers 7(67):2204–2215
6. Zolkov E, Weiss R, Cohen E (2020) Analysis and design of n-path band-pass filters with negative base band resistance. IEEE Trans Circuits Syst I Reg Papers 7(67):2250–2262
7. Zakharov A, Rozenko S, Pinchuk L, Litvintsev S (2021) Microstrip quazi-elliptic bandpass filter with two pairs of anti-parallel mixed-coupled SIRs. IEEE Microw. Wirel. Compon. Lett. 5(31):433–436

8. Zakharov A, Rozenko S, Zakharova N (2012) Microstrip bandpass filters on substrates with high permittivities. J Commun Technol Electron 3(57):342–351
9. Zakharov A, Il'chenko M (2010) A new approach to designing varicap-tuned filters. J Commun Technol Electron 12(55):1424–1431
10. Tsai H-J, Chen N-W, Jeng S-K (2013) Center frequency and bandwidth controllable microstrip bandpass filter design using loop-shaped dual-mode resonator. IEEE Trans Microw Theory Techn 10(61):3590–3600
11. Chu Q-X, Wang H (2008) A compact open-loop filter with mixed electric and magnetic coupling. IEEE Trans Microw Theory Techn 2(56):431–439
12. Yabuki H, Sagawa M, Matsuo M, Makimoto M (1996) Stripline dual-mode ring resonators and their application to microwave devices. IEEE Trans Microw Theory Techn 5(44):723–729
13. Makimoto M, Yamashita S (2001) Microwave resonators and filters for wireless communication.In: Theory, Design and Application. Springer, Heidelberg. https://doi.org/10.1007/978-3-662-04325-7
14. Zakharov A, Il'chenko M (2015) Pseudocombline bandpass filters based on half-wave resonators manufactured from sections of balanced striplines. J Commun Technol Electron 7(60):801–807
15. Zakharov A, Ilchenko M, Pinchuk L (2015) Coupling coefficient of quarter-wave resonators as a function of parameters of comb stripline filters. Radioelectron Commun Syst 6(58):284–289
16. Zakharov A, Il'chenko M (2013) Thin bandpass filters containing sections of symmetric strip transmission line. J Commun Technol Electron 7(58):728–736
17. Wolff I (1972) Microstrip bandpass filters using degenerate modes of a microstrip ring resonator. Electron Lett 12(8):302–303
18. Guglielmi M, Gatti G (1990) Experimental investigation of dual-mode microstnp ring resonators. European microwave conference 1990 Din., Budapest, Hungary, pp 901–906
19. Lei M-F, Wang H (2005) An analysis of miniaturized dual-mode bandpass filter structure using shunt-capacitance perturbation. IEEE Trans Microw Theory Techn 3(53):861–867
20. Chung M-S, Kim I-S, Yun S-W (2005) Varactor-tuned hairpin bandpass filter with an attenuation pole. Proceedings Asia-pacific microwave conference, Suzhou, China, December, volume 4, p 4
21. Zhang XY, Chen J-X, Xue Q, Li S-M (2007) Dual-band bandpass filters using stub-loaded resonators. IEEE Microw Wireless Compon Lett 8(17):583–585
22. Feng W, Che W (2012) Novel wideband differential bandpass filters. IEEE Trans Microw Theory Techn 6(60):1560–1568
23. Tsai H-J, Huang B-C, Chen N-W, Jeng S-K (2014) A reconfigurable bandpass filter based on a varactor-perturbed, T-shaped dual mode resonator. IEEE Microw. Wireless Compon. Lett. 5(24):297–299
24. Tang W, Hong J-S (2010) Varactor-tuned dual-mode bandpass filters. IEEE Trans Microw Theory Techn 8(58):2213–2219
25. Wang C-H, Lin Y-S, Chen CH (2004) Novel inductance-incorporated microstrip coupled-line bandpass filters with two attenuation poles. IEEE MTT-S international microwave symposium, Dig., pp 1979–1982
26. Wu C-H, Lin Y-S, Wang C-H, Chen C-H (2006) Novel microstrip coupled-line bandpass filters with shortened coupled sections for stopband extension. IEEE Trans Microw Theory Techn 2(54):540–546
27. Athukorala L, Budimir D (2009) Compact dual-mode open loop microstrip resonators and filters. IEEE Microw Wireless Compon Lett 11(19):698–700
28. Mao R-J, Tang X-H, Xiao F (2007) Compact quarter-wavelength resonator filter using lumped coupling elements. IEEE Microw Wireless Compon Lett 2(17):112–114
29. Zhang S, Zhu L, Weerasekera R (2015) Synthesis of inline mixed coupled quasi-elliptic bandpass filters based on λ/4 resonators. IEEE Trans Microw Theory Techn 10(63):3487–3493
30. Wu Y, Cui L, Zhuang Z, Liu Y (2018) A simple planar dual-band bandpass filter with multiple transmission poles and zeros. IEEE Trans Circuits Syst II Exp Briefs 1(65):56–60

31. Bi XK, Cheong P, Ho SK, Tam KW (2019) Design of dual-band bandpass filters with fixed and reconfigurable bandwidths based on terminated cross-shaped resonators. IEEE Trans Circuits Syst II Exp Briefs 3(66):317–321
32. Tang J, Liu H, Yang Y (2020) Balanced dual-band superconducting filter using stepped-impedance resonators with high band-to-band isolation and wide stopband. IEEE Trans Circuits Syst II Exp Briefs 1(68):131–135
33. Martín F, Zhu L, Hong J-S, Medina F (2018) Balanced Microwave Filters. Wiley, New York, NY, USA
34. Guo ZC, Zhu L, Wong SW (2019) A quantitative approach for direct synthesis of bandpass filters composed of transversal resonators. IEEE Trans Circuits Syst II Exp Briefs 4(66):577–581
35. Krishna I-S, Mukherjee S (2021) Triple-mode substrate integrated coaxial resonator based bandpass filter featuring flexible transmission zeros and adjustable bandwidth. IEEE Trans Circuits Syst II Exp Briefs 4(68):1223–1227
36. Lange DB (1964) Syntesis of transmission – line resonator with a specified frequency spectrum. Proc IRE 52(5):633–634
37. Litvinenko ON, Soshnikov VI (1964) Theory of nonuniform lines and their application in radio engineering. Sovetskoe Radio, Moscow (in Russian)
38. Richards PI (1948) Resistor-transmission-line circuits. Proc IRE 36(2):217–220
39. Riblet HJ (1957) General synthesis of quarter-wave impedance transformer. IRE Trans Microw Theory Techn MTT-5(1):36–43
40. Zakharov AV, Ilchenko MYe, Pinchuk LS (2014) Coupling coefficients of step-impedance resonators in stripline band-pass filters of array type. Radioelectron Commun Syst 5(57):217–223
41. Zakharov AV, Ilchenko MY, Karnauh VY, Pinchuk LS (2011) Stripline bandpass filters with step-impedance resonators. Radioelectron Commun Syst 3(54):163–169
42. Zakharov AV, Il'Chenko ME, Korpach VN (2014) Features of the coupling coefficients of planar stepped-impedance resonators at higher resonance frequencies and application of such resonators for suppression of spurious passbands. J Commun Technol Electron 6(59):550–556
43. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Trisection bandpass filter with mixed cross-coupling and different paths for signal propagation. IEEE Microw Wirel Compon Lett 1(30):12–15
44. Makimoto M, Yamashita S (1979) Compact bandpass filters using stepped impedance resonators. Proc IEEE 1(67):16–19
45. Makimoto M, Yamashita S (1980) Bandpass filters using parallel coupled stripline stepped impedance resonators. IEEE Trans Microw Theory Techn 12(28):1413–1417
46. Zakharov A (2021) Parametric and structural-parametric synthesis of nonuniform transmission line resonators. IEEE Trans Circuits Syst I Reg Papers 3(68):1055–1067
47. Shi J, Xue Q (2010) Dual-band and wide-stopband single-band balanced bandpass filters with high selectivity and common-mode suppression. IEEE Trans Microw Theory Techn 8(58):2204–2212
48. Zakharov A, Litvintsev S, Ilchenko M (2020) Transmission line tunable resonators with intersecting resonance regions. IEEE Trans Circuits Syst II Exp Briefs 4(67):660–664
49. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Hairpin resonators in varactor-tuned microstrip bandpass filter. IEEE Trans Circuits Syst II Exp Briefs 10(67):1874–1878
50. Guillemin EA (1959) Synthesis of passive networks: theory and methods appropriate to the realization and approximation problems. Wiley, New York
51. Matthaei GL, Young LT, Jones EM (1980) Microwave filters impedance-matching network and coupling structures. Artech House, Norwood, Massachusetts, US
52. Korn GA, Korn TM (1961) Mathematical handbook for scientists and engineers: definitions, theorems, and formulas for reference and review. McGraw-Hill, New York, NY, USA
53. Hong J-S (2011) Microstrip filters for RF/Microwave application, 2nd edn. Wiley, Hoboken, NJ, USA

54. Ilchenko MYu, Gorobets OYu, Bondar IA, Gaponov AM (2010) Influence of external magnetic field on the etching of a steel ball in an aqueous solution of nitric acid. J Magnetism Magnetic Mater 322(14):2075–2080
55. Ilchenko MYe, Kuzmin SYe, Narytnik TN, Belous OI, Radzikhovsky VN Transceiver for 130–134 GHz band digital radio relay system. Telecommun Radio Eng (English translation of Elektrosvyaz and Radiotekhnika) 72(17),1623–1638 (2013)

# Microwave Resonant Structures with Metamaterial Properties as Models of Some Quantum Interference Processes

**Hlib Avdieienko, Mykhailo Ilchenko, Roman Kamaraly, and Alexander Zhivkov**

**Abstract** Based on previously published models of microwave metamaterial structures simulating the Fano resonance, new models have been developed. Mathematical expressions have been obtained that describe characteristics such as electromagnetically Induced Transparency (EIT) and Autler-Tauns Splitting (ATS). The possibility of using the same models for simulating several resonant processes with a change in a number of their parameters is shown. The analysis was carried out both on the basis of the interaction of even and odd oscillations in microwave metamaterial structures, and on the basis of bridge resonant RLC quadrupoles based on lumped elements. Recommendations are given on the use of some special properties of the amplitude and frequency characteristics of microwave metamaterial structures to distinguish between processes of the EIT and ATS types.

**Keywords** Bridge equivalent circuits · Metamaterials · Fano resonance · Electromagnetically induced transparency · Autler-tauns splitting

## 1 Simulation of Electromagnetically Induced Transparency

Since the appearance in 1961 of one of the currently most cited scientific publications [1], interest in this phenomenon has been constantly growing. It is known that Fano type resonances are observed not only in quantum processes, but also in traditional

H. Avdieienko (✉) · M. Ilchenko (✉) · R. Kamaraly (✉) · A. Zhivkov (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Avenue 37, Kyiv 03056, Ukraine
e-mail: django2006@ukr.net

M. Ilchenko
e-mail: ilch@kpi.ua

R. Kamaraly
e-mail: theromarioagro@gmail.com

A. Zhivkov
e-mail: a.zhivkov@kpi.ua

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_25

microwave electrodynamics and optics [2], photonics [3], and acoustics [4]. It was shown in [5] that lumped-element bridge RLC quadrupoles can be used to effectively model interference processes of the Fano resonance type. It is noted in [3] that electromagnetically induced transparency (EIT) can be viewed as a special case of Fano resonance when the frequencies of strongly and weakly damped oscillators match, i.e. $\omega_1 = \omega_2$.

Electromagnetically Induced Transparency can be considered as a specific case of Fano resonance when the frequencies of interfering oscillations coincide. Therefore, it seems appropriate to extend the approaches used in [5] to the analysis of processes, as noted in [6], for EIT simulation.

Based on the general formulas obtained in [7] for an eight-terminal network consisting of two waveguides connected along a wide wall with holes in the elliptical polarization region and dielectric resonators located in them, the transfer and reflection coefficients (T and R are equal to the $S_{21}$ and $S_{11}$ parameters of the matrix scattering) were presented in [5] for a particular case of an eight-pole - a four-pole in the form of a bridge bandstop filter and expressed as

$$\mathrm{T} = 1 - K_1/(1 + K_1) - K_2/(1 + K_2), \tag{1}$$

$$\mathrm{R} = K_2/(1 + K_2) - K_1/(1 + K_1), \tag{2}$$

where $\mathrm{K}_1$, $\mathrm{K}_2$ – coupling coefficients of the first and second resonators with the transmission line:

$$K_1 = \mathrm{K}_1/[(1 + \mathrm{j}(\xi + a)], \tag{3}$$

$$K_2 = \mathrm{K}_2/[(1 + \mathrm{j}(\xi - a) \cdot b], \tag{4}$$

$$\mathrm{K}_1 = Q_{\mathrm{ext1}}/Q_{01}, \tag{5}$$

$$\mathrm{K}_2 = Q_{\mathrm{ext2}}/Q_{02}, \tag{6}$$

where $Q_{0i}$ is own quality factor of $i$-resonator; $Q_{\mathrm{ext}i}$ is loaded quality factor of $i$-resonator, $i = 1,2$; $\xi$ is generalized detuning of frequency with respect to the central frequency $f_0$ of the filter; and $a$ is generalized frequency detuning of "magnetic" $f_m$ and "electric" $f_e$ oscillations with respect to $f_0$

$$f_0 = (f_m + f_e)/2, \tag{7}$$

$$a = 2(f_e - f_m)/(f_e + f_m) \times \sqrt{Q_{01} * Q_{02}}, \tag{8}$$

$$\xi = 2(f - f_0)/(f + f_0) \times \sqrt{Q_{01} * Q_{02}}, \tag{9}$$

$$b = Q_{01}/Q_{02}. \tag{10}$$

It is important to note that the use of the terms "electric" and magnetic in relation to the types of oscillations is correct if the resonators are modeled by electric and magnetic dipoles. It can also be in-phase and anti-phase (symmetric and antisymmetric) [8], even and odd [9] types of oscillations, as well as "oscillator" and "rotator" depending on their phase portrait [10]. Since further processes are modeled using parallel and series resonators in the arms of a bridge quadrupole, we will designate resonant frequencies of parallel and series resonators as $f_p$ and $f_s$ accordingly.

Considering that EIT occurs when the frequencies of interfering oscillations are equal (detuning a = 0), let us consider the frequency dependences (in relative detunings ξ) of the transmission coefficient T (Eq. (1)) and the Group Delay Time (GD) of individual oscillations and the filter as a whole. The gains (green and dotted red curves) of individual resonators and the entire filter (blue curve) for two different values of $b$ are shown in Fig. 1a, b. Case $b = 1$ corresponds to the variant with resonators of equal quality factor, In case when $b = 2.6$, the quality factor of the first, less loaded resonator ($K_1 = 1$, $K_2 = 30$) is 2.6 times higher than the quality factor of the second, heavily loaded resonator.

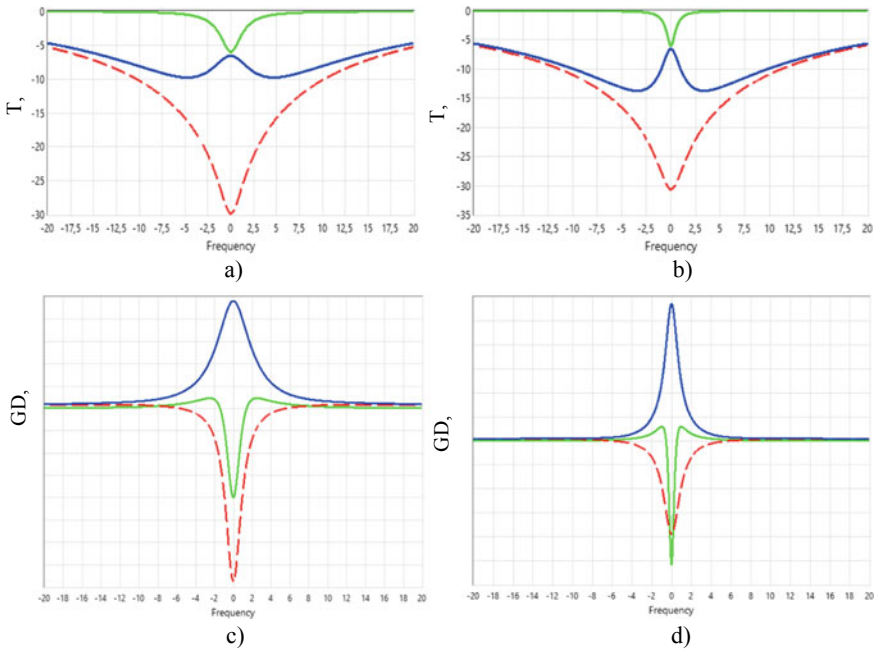



**Fig. 1** The gains **a**, **b** and group delay **c**, **d** (green and dotted red curves) of individual resonators and the entire filter (blue curve) for two different values of parameter $b$: 1) $K_1 = 1$, $K_2 = 30$, $b = 1$ **a**, **c**; 2) $K_1 = 1$, $K_2 = 30$, $b = 2.6$ **b**, **d**

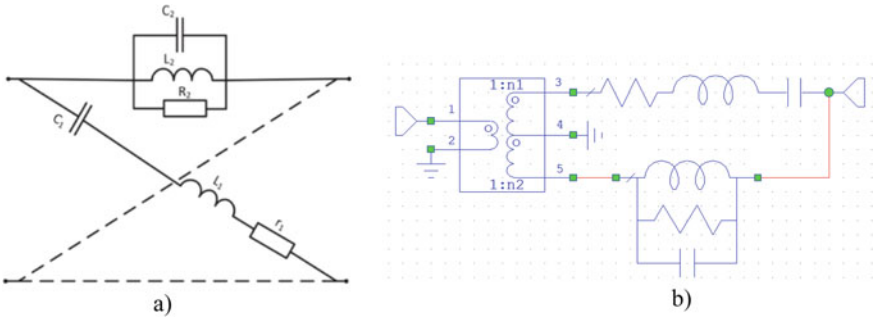



**Fig. 2** Equivalent circuit of a lumped-element bridge bandstop filter used to simulate the Fano **a** and bridge circuit based on the ideal symmetrical mid-point transformers **b**

The GD characteristics of the individual resonators and the filter as a whole (the colors of the GD curves correspond to the colors of the gains T) are presented in Fig. 1c, d. Since GD was not calculated by formulas but was calculated as a derivative of the corresponding phase characteristic, the ordinate axes of the graphs in Fig. 1c, d are dimensionless, and the ratios between "multi-colored GDs" can only be established within limits of individual graphs. Nevertheless, we will return to this problem after analyzing the characteristics of bridge bandstop filters with parallel and series oscillatory circuits on lumped elements using the AWR program.

Figure 2a shows the equivalent circuit of a lumped-element bridge bandstop filter used to simulate the Fano resonance in paper [11]. However, it cannot be directly used for calculations in the AWR program, in which bridge circuits are presented in the form of Fig. 2b using ideal symmetrical mid-point transformers. The equivalence of these circuits was demonstrated in [12], the ratios of the reactive resistances of the arms for different versions of the bridges are also presented in [13].

Electrical circuits with resonators based on lumped elements in the form of coupled circuits for modeling resonant quantum effects have been used for a long time [14–16]. At the same time, it is known that it is bridge quadrupoles that are the most common schemes of symmetrical circuits (compared to T- and U-shaped quadrupoles) [17]. It is also significant that bridge quadrupoles with certain ratios of the arm parameters are characterized by the non-minimum phase property [18], which, as shown in [5], makes them indispensable in the analysis of Fano resonance-type processes.

Further analysis of resonance processes will be carried out in the range of centimeter wavelengths. This will enable us at least qualitatively compare our results with the results of simulation of similar processes using metamaterials discussed above. If it is necessary to simulate at lower frequencies, the parameters of resonant circuits on lumped elements can be easily scaled. Therefore, the question of the physical "feasibility" of certain values of inductances or capacitances of parallel or series resonant circuits is not fundamental here.

Figure 3a, b shows the characteristics of the transfer coefficients and the corresponding delay parameters (GD) of quadrupoles (individual arms are red and green

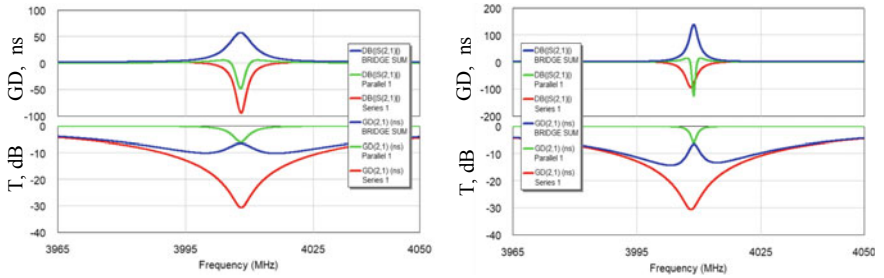



**Fig. 3** Characteristics of the transfer coefficients (T) and the corresponding delay parameters (GD) of quadrupoles simulated in AWR: **a** $K_1 = 1$, $K_2 = 30$, b = 1; **b** $K_1 = 1$, $K_2 = 30$, b = 2.6

curves, quadrupoles are blue curves) with arm transfer coefficients similar to the transfer coefficients calculated by formula (1) and shown in Fig. 1a–Fig. 1d). In this case, the coupling coefficients are calculated as

$$K_S = 2R_L/R_S, \tag{11}$$

$$K_P = R_P/2R_L, \tag{12}$$

where $R_L$ is the load resistance of the four-terminal network (in our case the typical value $R_L = 50\ \Omega$ was used - that is, the resistance of the input and output transmission lines), $R_S$ and $R_P$ are the resistances of the series and parallel circuits at resonant frequencies.

The excellent correspondence of the characteristics of the quadrupoles to the characteristics calculated by the formulas indicates that expression (1) can be used for a detailed analytical study of the possibility of achieving large values of GD with relatively small losses of the quadrupole. This fact is the reason for the increased attention of researchers to the EIT phenomenon [19–22].

Note that the results of EIT simulation for metamaterials presented, for example, in [23–25], are significantly more modest than those stated in [21]. At the same time, as can be seen from Fig. 1d and Fig. 2b, at the same attenuation levels provided by the higher quality resonator included in the quadrupole (green curves T and GD), the quadrupole as a whole (blue lines T and GD) provides the same level of delay (GD), but with the opposite sign. At the same time, which is very significant, this delay is much broader in comparison to delay of a high-Q resonator. It rather corresponds in terms of broadband to the bandwidth of the second, strongly coupled, low-Q resonator (red lines T and GD).

Figure 4 shows the characteristics of the same bridge quadrupole under Fano resonance conditions, achieved by tuning a strongly coupled resonator (red curve in the figures) down and up in frequency relative to the EIT mode.

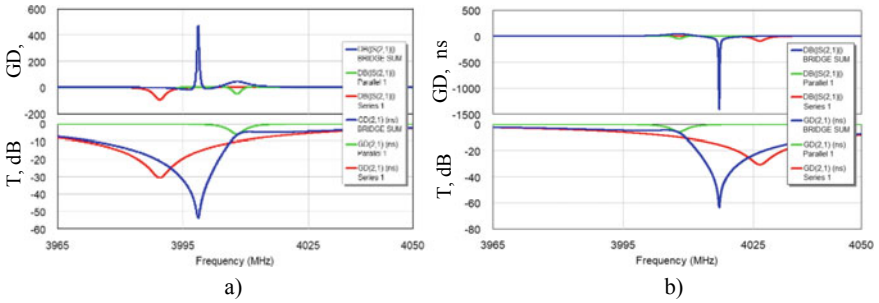



**Fig. 4** Characteristics of the transfer coefficients (T) and the corresponding delay parameters (GD) of quadrupoles achieved by tuning a strongly coupled resonator (red curve in the figures) down **a** and up in frequency **b** relative to the EIT mode

As one can see that in Fano resonance modes extremely large delay levels are achievable (theoretically its strives to infinity; in practice, the delay is limited by the quality factor of the resonators, that is, the signal levels comparable to the noise levels in actually used physical materials). However, the levels of the signals themselves are very small. Figure 5a shows the model of an oscillation intermediate between the Fano resonance and the EIT, obtained using the same bridge quadrupole. It is worth noting that similar pictures for the EIT are presented in [26].

Similar gain characteristics of T parameter obtained by using formula (1) are presented in Fig. 5b. They are presented solely to demonstrate the effectiveness of the proposed simulation method in AWR software package.

To study effects similar to Fano resonance and EIT, a variety of structures are used as unit cells of metamaterials, such as paired ring resonators (PRRs) [8], split-ring resonators (SRR) and two-gap SRR [15], bilayered fish-scale metamaterial and its unit cell [23] and various other structures [1, 2]. It seems that it is most convenient for numerical and experimental simulation to use a structure in the form of a bandstop filter (Fig. 6a), consisting of two resonators of different lengths. The ratio of the lengths of the short and long resonators for such filter is about 1:2. The metamaterial

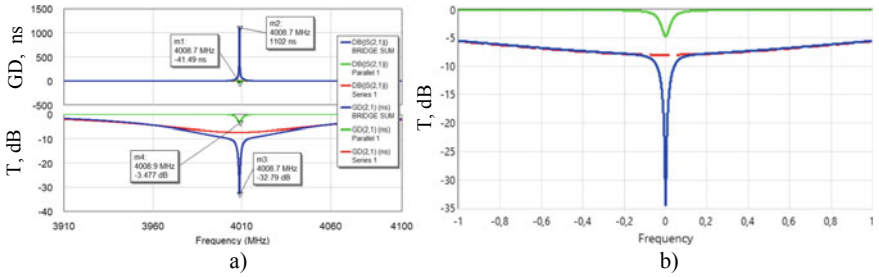



**Fig. 5** **a** Simulation results of transfer and group delay characteristics for oscillation intermediate between the Fano resonance and the EIT; **b** calculation results of transfer characteristics for oscillation intermediate between the Fano resonance and the EIT

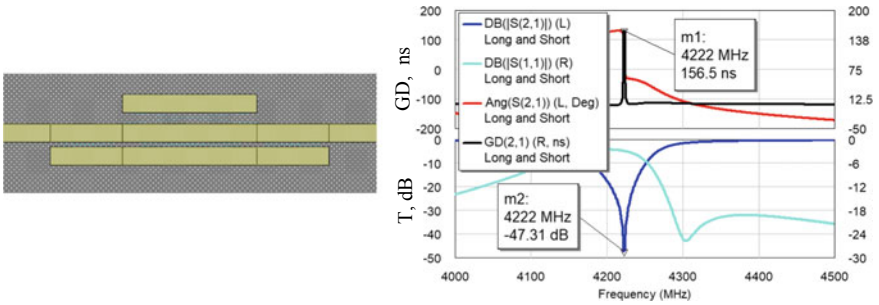



**Fig. 6** **a** Metamaterial structure in the form of a two-resonator microwave bandstop filter; **b** Simulation results of transfer (blue line), reflection (white blue), phase (red line) and group delay (black line) characteristics of microwave bandstop filter

properties of this structure, first described in [27], were declared and then repeatedly confirmed in [5, 6, 13, 28, 29, 31]. Indeed, the resonant frequencies of individual resonators are easily controlled by their lengths, and the connection with the line (in other words, the loaded Q-factors and coupling coefficients from (5) and (6)) is easily controlled by the size of the gaps between the line and the resonators.

For characteristics calculating in the AWR software package, a dielectric substrate ($\varepsilon = 10$, $\mathrm{tg}\delta = 0.002$) with 2 mm thickness was used. The calculated attenuation of –47 dB was achieved with resonator lengths of 14 and 29.5 mm and gaps between the resonators and the line of 0.5 and 1.125 mm accordingly. A short resonator simulates an odd and a long resonator simulates an even oscillation. The results of numerical (software-topological) simulation in the AWR software package are shown in Fig. 6b.

Experimental simulation (Fig. 7a, b) did not allow us, unfortunately, to achieve the same "impressive" results in GD values. This is due to the difficulty of fine tuning the gaps and lengths. However, it demonstrated obvious "metamaterial" properties of the cell in the form of an anomalously large attenuation (~50 dB), a steep section of the phase characteristic (phase changes by 70° when the frequency changes by 19 MHz) and a large positive value of group delay (~60 ns, with group delay of separate resonators detuned in frequency no more than –10 ns).

The considered model of the bandstop filter makes it possible to "achieve" attenuation values of the order of 100 dB and GD of about 50 μs when setting a high quality factor of the material and accuracy of calculations. However, in practice, such parameters are unattainable due to influence of the real quality factor, manufacturing process and tuning accuracy and are not considered here.

If significantly (up to 6 mm) increase the gap between the transmission line and the short resonator in comparison with the case of calculations for the Fano resonance, it turns out that aforementioned filter model will demonstrate characteristics of the EIT type. There will be a not very deep, but obvious "deep" in the cut-off band ("transparency window"), as well as a positive group delay value. The obtained

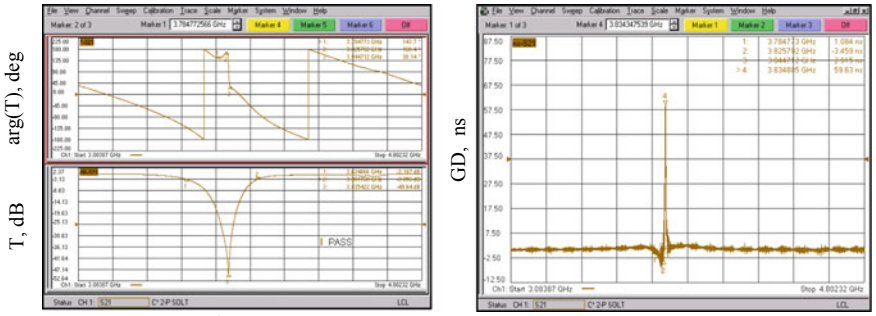



**Fig. 7** Results of the experimental simulation of metamaterial structure in the form of a two-resonator microwave bandstop filter: **a** transfer and phase characteristics; **b** characteristic of the group delay

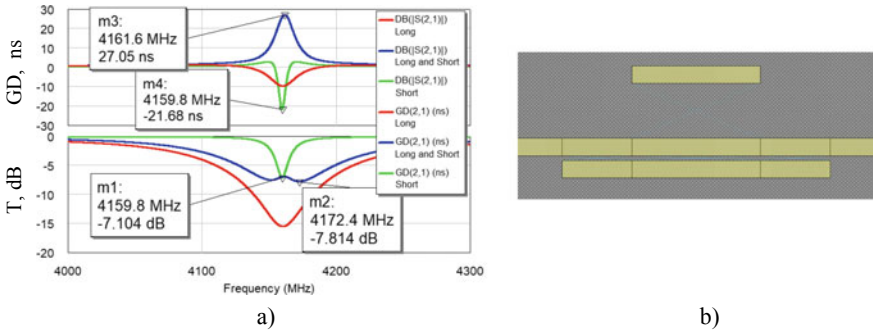



**Fig. 8** **a** Simulation results of transfer and group delay characteristics of microwave bandstop filter; **b** Metamaterial structure in the form of a two-resonator microwave bandstop filter for the case when gap between the transmission line and the short resonator is 6 mm

characteristics are presented in Fig. 8a. Also Fig. 8b demonstrates the topology of our filter with new gaps.

As with the use of the previously considered EIT models, we see that with comparable levels of attenuation in the transparency band and attenuation of higher quality weakly coupled resonators, the levels of delays achieved are also approximately the same, but the spectral response of the delay of the entire quadrupole is much wider.

## 2 Modeling the Stark Effect

After successfully simulation of the EIT using a bandpass filter (Fig. 8b), let's pay attention to its characteristics in a wider frequency range (Fig. 9a).

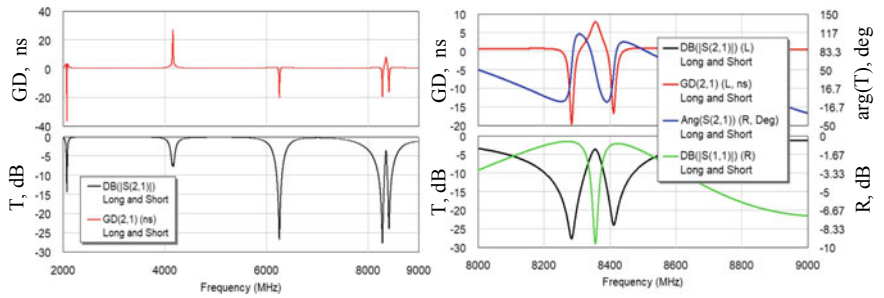



**Fig. 9** Simulation results of microwave bandstop filter transfer, phase and group delay characteristics in 2–9 GHz frequency band **a** and more detailed in 8–9 GHz band **b**

The filter gains (black curves) and their corresponding GD parameters (red curves) can be clearly classified into types:

1. Curves in the region of 2 GHz—the first harmonic of a long resonator and normal GD corresponding to one resonance.
2. Curves in the region of 4 GHz—the interference curve of the 1st and 2nd harmonics of the short and long resonators, respectively, just considered above - the EIT characteristic. This follows at least from a sufficiently significant positive value of GD.
3. Curves around 6 GHz—the third harmonic of a long resonator. Sufficiently large attenuation due to the strong coupling of the extended resonator with the line, negative, but not very large value of GD due to the low quality factor of this oscillation (radiation losses increase).
4. Finally, the curves in the region of 8 GHz are the result of the interference of two even oscillations - the 2nd harmonic of the short resonator and the 4th harmonic of the long resonator. These characteristics, which are the result of the interference of two closely spaced, but not coinciding in frequency, vibrations of the same type, are commonly referred to as the so-called Autler-Townes Splitting (ATS) [30].

The manifestation of ATS is primarily a specific splitting of spectral lines in optics [26]. The differences between EIT and ATS have been studied sufficient detailed [32, 33]. In paper [14], characteristics similar to ATS are modeled using coupled RLC circuits. In this work, we would like to demonstrate the possibility of modeling ATS-like effects using the simplest metamaterial cells and lumped-element RLC bridge structures. These methods turn out to be quite efficient even though the ATS effect is not a metamaterial one and the microwave structures used for simulation do not show their metamaterial properties. So, a detailed "portrait" of the characteristics of the investigated quadrupole (reflection coefficient $S_{11}$ and the phase of the gain $S_{21}$) in the range of 8–9 GHz is shown in Fig. 9b. A similar picture can be obtained if we consider the interaction of the 1st harmonic of a short resonator and the 3rd harmonic of a long one in the same structure of a microstrip bandstop filter (Fig. 10a).

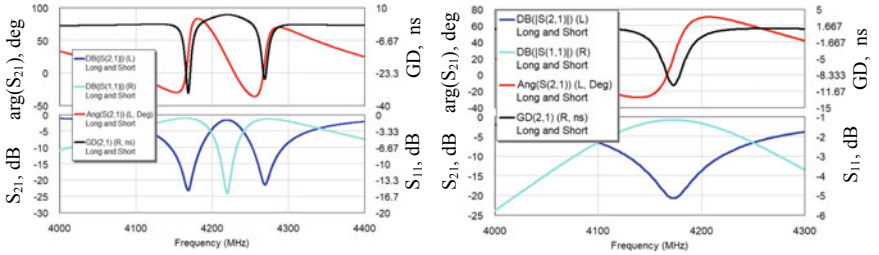



**Fig. 10** Simulation results of microwave bandstop filter transfer, phase and group delay characteristics in 4–4.4 GHz band for the case of interaction of the 1st harmonic of a short resonator and the 3rd harmonic of a long one

From what has been said, it follows that effects similar to ATS arise in the microwave frequency range as a result of the interference of oscillations of the same "parity" which close in frequency to each other—"even" or "odd". If the resonant frequency of the 1st oscillations is slightly changed (by lengthening or shortening the corresponding resonator in the model), then the splitting disappears and the total oscillation of two resonators with the same frequency becomes similar to the oscillation of single resonator (Fig. 10b). Accordingly, the equivalent circuit of a such process using resonators based on lumped RLC elements can be represented as parallel-connected series circuits or series-connected parallel oscillatory circuits (Fig. 11).

Strictly speaking, the use of a bridge circuit for modeling ATS is not necessary. One can get by with one of its arms, but we will need the "bridge" later when we demonstrate the transition states between EIT and ATS. Figure 12 shows the characteristics of the quadrupole presented in Fig. 11a for different values of the capacitances of the series circuits.

ATS variants are considered in Fig. 13 at different resonant frequencies and coupling coefficients of series circuits with a source and a load. As expected, with an increase in the mutual detuning of the resonators, their amplitude and phase characteristics become almost equivalent to the characteristics of individual series resonant circuits and the large GD value of one of the circuits in Fig. 13b is due to its very

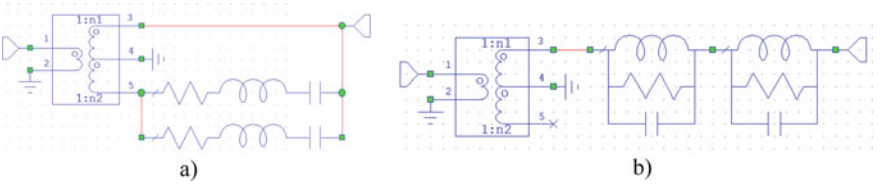



**Fig. 11** Equivalent circuit with resonators based on lumped RLC elements for ATS process simulation

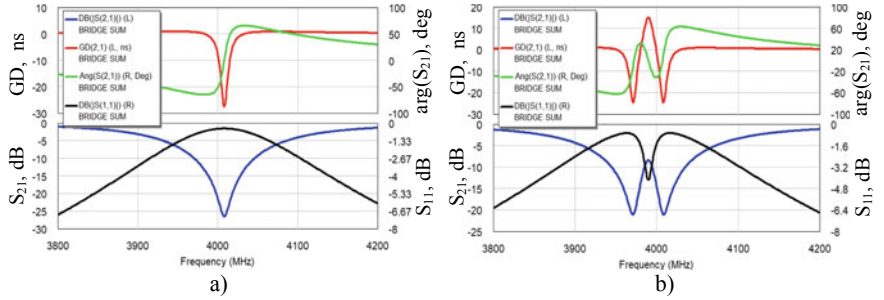



**Fig. 12** Characteristics of the quadrupole presented in Fig. 11a for the case $L_1 = L_2 = 146$ nH, $R_1 = R_2 = 10\ \Omega$: **a** $C_1 = C_2 = 0.0108$ pF; **b** C2 = 0.011 pF. Source and load resistances are the same and equal to $R_L = 50\ \Omega$

strong connection ($K_{S2} = 100$ in expression (11)) with the circuit input and output and a large intrinsic quality factor $Q_0 \sim 3500$ [34].

Figure 14 shows the characteristics of the bridge circuit presented in Fig. 11b, also simulating ATS-type characteristics. As one can see, they are practically equivalent to the characteristics obtained using the circuit in Fig. 11a, so we will not consider them further in more detail.

It is much more interesting to consider the diagram of a bridge quadrupole in Fig. 15a. Its characteristics are presented in Fig. 15b. The resonant frequencies of the first parallel (ID = PRLC1) and serial (ID = PRLC3) coincide, the oscillations are "subtracted", their total loss in the resonant frequency range ~3790 MHz is only 1.3 dB. The total GD ~9 ns is positive, which also indicates the mode EIT in this frequency range. By rearranging the first parallel circuit upward (reducing C_Par1 to 26.1 pF), we achieve Fano resonance at 3871.2 MHz, with anomalously large attenuation and GD values (Fig. 15c). Finally, when C_Par1 is further reduced to 25 pF, an interference pattern of the ATS type appears between two resonances of parallel circuits that are already close in frequency (Fig. 15d). Thus, the bridge

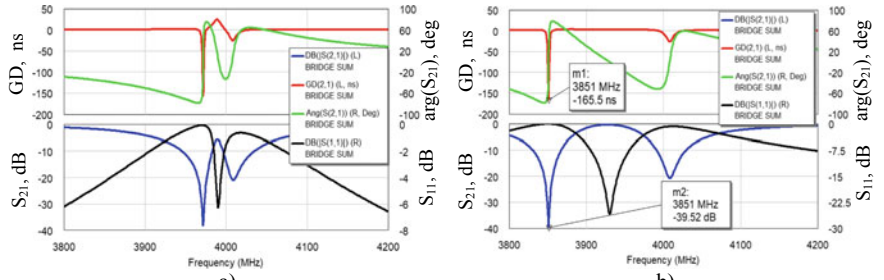



**Fig. 13** Characteristics of the quadrupole presented in Fig. 11a for the case $L_1 = L_2 = 146$ nH, $R_1 = 10\ \Omega$, $R_2 = 1\ \Omega$: **a** $C_1 = 0.0108$ pF, $C_2 = 0.011$ pF; **b** C2 = 0.0117 pF. Source and load resistances are the same and equal to $R_L = 50\ \Omega$

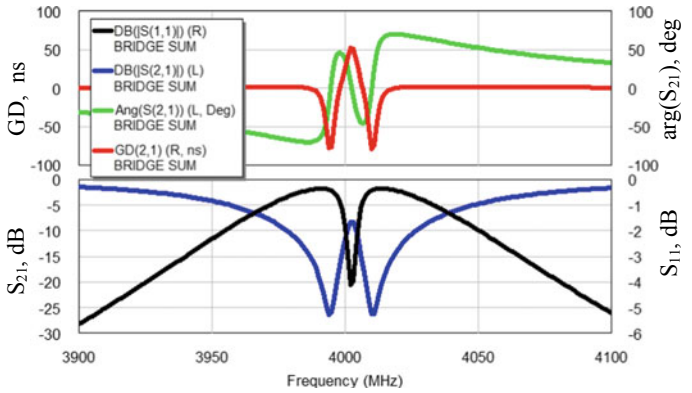



**Fig. 14** Characteristics of the quadrupole presented in Fig. 11b for the case $L_1 = 0.0635$ nH, $L_2 = 0{,}063$ nH, $R_1 = R_2 = 2000\ \Omega$, $C_1 = C_2 = 25$ pF. Source and load resistances are the same and equal to $R_L = 50\ \Omega$

circuit of the quadrupole presented in Fig. 15a allows you to "trace" the evolution of interference processes from EIT to ATS, as was previously for the case with the scheme of microstrip bandstop filter (Fig. 6a).

For the case of interaction (interference) of oscillations of the same type (even or odd), expressions (1)–(2) are converted to the form:

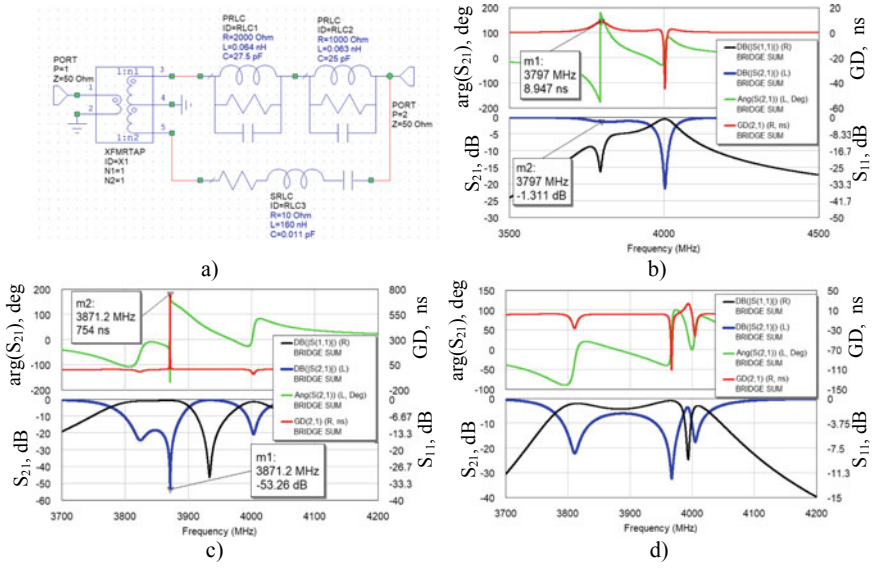



**Fig. 15** **a** Diagram of a bridge quadrupole; **b** characteristics for C_Par1 = 27.5 pF, EIT mode at 3797 MHz - two oscillations compensate each other and create a window of transparency; **c** Fano resonance mode at 3871.2 MHz, C_Par1 = 26.1 pF; **d** ATS mode at 4 GHz, C_Par1 = 25 pF

$$\mathbf{T} = 1 - K_{sum}/(1 + K_{sum}) \tag{13}$$

$$\mathbf{R} = K_{sum}/(1 + K_{sum}) \tag{14}$$

where for even type of oscillations

$$K_{sum} = K_{even1} + K_{even2}, \tag{15}$$

and respectively for odd type of oscillations

$$K_{sum} = K_{odd1} + K_{odd2}, \tag{16}$$

$K_{even1}$, $K_{even2}$, $K_{odd1}$, $K_{odd2}$ are coupling coefficients of resonators (oscillations) with the transmission lines, determined by expressions (3)–(10).

Based on expressions (13)–(14) for equal coupling coefficients ($K_1 = K_2 = 5$) and a slight detuning between the resonators ($a = 6$), the dependences presented in Fig. 16 were calculated. They are completely similar in shape to the previously considered ATS characteristics of microstrip bandstop filter in Fig. 10a and the bridge filter formed by the parallel inclusion of two mutually disordered series resonators (Fig. 12b).

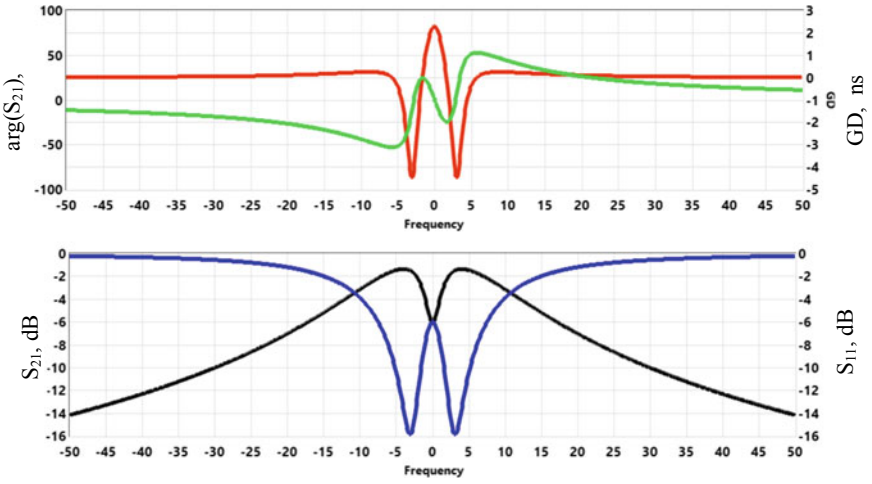



**Fig. 16** Characteristics according to formulas (13)–(14): $K_1 = K_2 = 5$, $a = 6$

## 3 Fundamental Differences Between EIT and ATS

Since splitting (degeneracy removal of two or more oscillations) of the ATS and EIT type occurs in calculations and experiments quite often, the question of their correct identification is still relevant not only in electrodynamics but also in acoustics [14, 25, 26, 35]. We want to note some very principled positions that distinguish these types of splitting:

- Completely different behavior of phase characteristics (two bends) and, as a consequence, group delay.

Phase characteristics and their derivatives—GD, considered earlier and presented together for ease of comparison in Fig. 17 indicate the presence of two "bends" of the ATS phase and one, rather steep (and corresponding to the area of the "negative" dispersion, as evidenced by the positive value of GD) for EIT. As a consequence, there are negative peaks on the both sides of the positive peak of GD for ATS and no peaks of GD for EIT. The amplitude characteristics of ATS and EIT are similar in shape and almost no suitable for identification.

Figure 18 shows fragments of "emissions" that are formed for the transmission coefficient T of the total quadrupole due to the interaction of low-Q "red" and high-Q "green" oscillations.

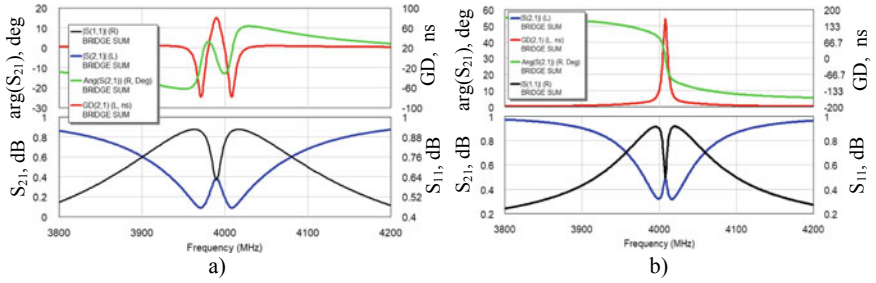



**Fig. 17** Comparison of ATS and EIT: **a** Characteristics for ATS; **b** Characteristics EIT

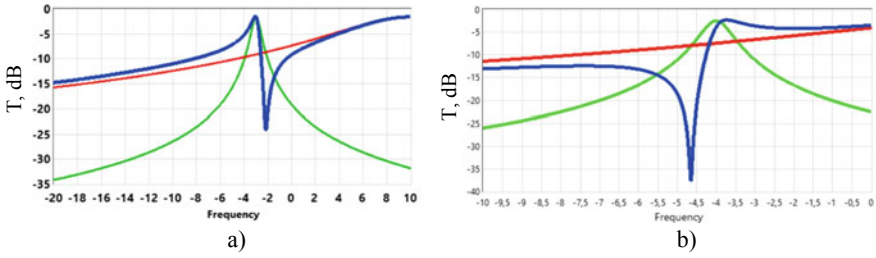



**Fig. 18** Transmission coefficient T for: **a** ATS, **b** EIT

In this case, splitting types (ATS or EIT) can also be effectively identified by amplitude characteristics, and, as can be judged from the pictures from [36], the splitting considered there belongs precisely to the ATS type, as well as the splitting in a Z-like structure considered in [37].

In this case, splitting types can also be effectively identified by amplitude characteristics, and, as can be judged from the pictures from [36], the splitting considered there in a Z-like structure belongs specifically to the ATS type.

Since, as noted in [35], the characteristics of splitting of the ATS and EIT types in electrodynamics and acoustics are similar, and in [37] the possibility of modeling an acoustic Helmholtz resonator in the mode of excitation by a bridge circuit is declared, the bridge models and formulas which were considered can also be used for similar research in acoustics.

## 4 Conclusions

Calculations of Fano resonances, splittings of the EIT and ATS types were carried out by various methods. The results coincide with each other and with experiments, one's own and others', therefore, the proposed modeling methods are adequate and can be useful in the analysis and prediction of some quantum processes. It is essential that there is no need to be limited to any special types of vibrations. The results obtained can be extended to any so-called "normal" oscillations, and not necessarily coupled oscillations.

The presence of simple analytical expressions that are adequate to physical models makes it possible, if necessary, to find quite effectively the optimal relationships between the physical parameters of a metamaterial cell in generally accepted terms of resonant frequencies, natural and loaded Q factors, comparing them with similar parameters of quantum processes (spectrum width, energy, the same resonant frequencies etc.). Elementary cells of metamaterials in the form of microstrip bandstop filters are easy to manufacture and, therefore, can be effectively used in the physical laboratories of universities to simulate the quantum resonance effects considered above.

**Acknowledgements** The present work was supported by the National research foundation of Ukraine, project "Microwave devices based on resonant structures with metamaterial properties for the life protection and information security of Ukraine" (ID 2021.01/0030).

## References

1. Fano U (1961) Effects of configuration interaction on intensities and phase shifts. Phys Rev 124(6):1866–1878. https://doi.org/10.1103/physrev.124.1866

2. Kamenetskii E, Sadreev A, Miroshnichenko A (2018) Fano Resonances in Optics and Microwaves. Springer, Heidelberg. https://doi.org/10.1007/978-3-319-99731-5
3. Limonov M, Rybin M, Poddubny A et al (2017) Fano resonances in photonics. Nat Photon 11:543–554. https://doi.org/10.1038/nphoton.2017.142
4. Hein S, Koch W, Nannen L (2010) Fano resonances in acoustics. J Fluid Mech 664:238–264. https://doi.org/10.1017/s0022112010003757
5. Ilchenko M, Zhivkov A (2019) Bridge equivalent circuits for microwave filters and fano resonance. In: Ilchenko M, Uryvsky L, Globa L (eds) Advances in Information and Communication Technologies. UKRMICO 2018. LNEE, vol 560, pp 278–298. Springer, Cham. https://doi.org/10.1007/978-3-030-16770-7_14
6. Ilchenko M, Zhivkov A, Zazerin A (2018) Metamaterial cells equivalent circuit, degenerated oscillation and electromagnetically induced transparency. In: 2018 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo). https://doi.org/10.1109/ukrmico43733.2018.9047596
7. Ilchenko ME, Zhivkov AP (1989) Microwave device with multiple modes of dielectric resonators. Izvestija Vysshih Uchebnyh Zavedenij. Radioelektronika 32(5):56–59
8. Liu Y, Fang N, Wu D, Sun C, Zhang X (2007) Symmetric and antisymmetric modes of electromagnetic resonators. Appl Phys A 87(2):171–174. https://doi.org/10.1007/s00339-006-3837-0
9. Chen WH (1970) Even and odd mode impedance of coupled pairs of microstrip lines (correspondence). IEEE Trans Microw Theory Tech 18(1):55–57. https://doi.org/10.1109/tmtt.1970.1127138
10. Sommerfeld A (1956) Atomic Structure and Spectral Lines, vol 1. State Publishing House of Technical Literature, Moscow
11. Zhivkov AP, Shevtsov KO, Kamarali RV, Orlov AT (2019) Fano resonance and metamaterial cells equivalent circuits. In: 2019 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo). https://doi.org/10.1109/ukrmico47782.2019.9165428
12. Cauer W (1958) Synthesis of Linear Communication Networks. McGraw-Hill, New York
13. Ilchenko ME, Zhivkov AP, Orlov AT (2016) Filters based on resonators with modes similar in frequency as cells of metamaterials. Naukovi Visti NTUU KPI 104(1):7–14. (in Russian)
14. Garrido Alzar CL, Martinez MAG, Nussenzveig P (2002) Classical analog of electromagnetically induced transparency. Am J Phys 70(1):37–41. https://doi.org/10.1119/1.1412644
15. Tassin P, Zhang L, Koschny T, Economou EN, Soukoulis CM (2009) Planar designs for electromagnetically induced transparency in metamaterials. Opt Express 17(7):5595. https://doi.org/10.1364/oe.17.005595
16. Harden J, Joshi A, Serna JD (2011) Demonstration of double EIT using coupled harmonic oscillators and RLC circuits. Eur J Phys 32(2):541–558. https://doi.org/10.1088/0143-0807/32/2/025
17. Ernst A (1957) Guillemin Synthesis of Passive Networks: Theory and Methods Appropriate to the Realization and Approximation Problems. John Wiley & Sons, Hoboken
18. Siebert WM (1985) Circuits, Signals, and Systems. The MIT Press, Cambridge
19. Harris SE (1997) Electromagnetically induced transparency. Phys Today 50:36–42
20. Boller KJ, Imamoglu A, Harris SE (1991) Observation of electromagnetically induced transparency. Phys Rev Lett 66:2593–2596
21. Hau LV, Harris SE, Dutton Z, Behroozi CH (1999) 10 light speed reduction to 17 m/s in an ultracold atomic gas. Nature 397:594–598
22. Aleksandrov EB, Zapasskii VS (2006) Physics-Uspekhi 49(10):1067. https://doi.org/10.1070/pu2006v049n10abeh006056
23. Papasimakis N, Fedotov VA, Zheludev NI, Prosvirnin SL (2008) Metamaterial analog of electromagnetically induced transparency. Phys Rev Lett 101(25). https://doi.org/10.1103/physrevlett.101.253903

24. Nakanishi T, Kitano M (2018) Storage and retrieval of electromagnetic waves using electromagnetically induced transparency in a nonlinear metamaterial. Appl Phys Lett 112(20):201905. https://doi.org/10.1063/1.5035442
25. Malik J, Oruganti SK, Song S, Ko NY, Bien F (2018) Electromagnetically induced transparency in sinusoidal modulated ring resonator. Appl Phys Lett 112(23):234102. https://doi.org/10.1063/1.5029307
26. Peng B, Özdemir ŞK, Chen W, Nori F, Yang L (2014) What is and what is not electromagnetically induced transparency in whispering-gallery microcavities. Nat Commun 5(1). https://doi.org/10.1038/ncomms6082
27. USSR Inventor's Certificate 1529321
28. Ilchenko ME, Zhivkov AP (2017) Areas of degeneration oscillations in metamaterial cells. In: 2017 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo). https://doi.org/10.1109/UkrMiCo.2017.8095389
29. Ilchenko M, Kamarali R, Zhivkov A, Kopaniev M, Saichenko I (2020) Non-Lorentzian resonance characteristics of metamaterial cells in a waveguide. In: 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), 2020, pp 543–546. https://doi.org/10.1109/PICST51311.2020.9467961
30. Autler SH, Townes CH (1955) Stark effect in rapidly varying fields. Phys Rev 100(2):703–722. https://doi.org/10.1103/physrev.100.70
31. Wei B, Jian S (2017) Objectively discriminating the optical analogy of electromagnetically induced transparency from Autler-Townes splitting in a side coupled graphene-based waveguide system. J Opt 19(11):115001. https://doi.org/10.1088/2040-8986/aa8c56
32. Abi-Salloum TY (2010) Electromagnetically induced transparency and autler-townes splitting: two similar but distinct phenomena in two categories of three-level atomic systems. Phys Rev A 81(5). https://doi.org/10.1103/physreva.81.053836
33. He L-Y, Wang T-J, Gao Y-P, Cao C, Wang C (2015) Discerning electromagnetically induced transparency from Autler-Townes splitting in plasmonic waveguide and coupled resonators system. Opt Express 23(18):23817. https://doi.org/10.1364/oe.23.023817
34. Pozar D (2011) Microwave Engineering, 4th edn. Wiley, Hoboken, 22 November 2011
35. Cheng Y, Jin Y, Zhou Y, Hao T, Li Y (2019) Distinction of acoustically induced transparency and Autler-Townes splitting by Helmholtz resonators. Phys Rev Appl 12(4). https://doi.org/10.1103/physrevapplied.12.044025
36. Bochkova E, Shah NB, De André L, Lupu A (2015) Engineering dark mode resonances in Z-metasurfaces for sensing applications. In: First International Workshop on Metamaterials-by-Design, December 2015. https://www.researchgate.net/publication/305650267
37. Haiko H, Zhivkov O, Pyha L (2021) Application of resonant oscillatory systems for the seafloor gas hydrates development. In: E3S Web of Conferences, Gas Hydrate Technologies: Global Trends, Challenges and Horizons–2020, vol 230, p 01020. https://doi.org/10.1051/e3sconf/202123001020

# Modification of Capon's Method for Several Radio Sources Coordinates Determining by the Shape of the Electromagnetic Wave Phase Front

**Hlib Avdieienko and Yevhenii Yakornov**

**Abstract** Inefficiency of classical Capon's method of bearing angles estimation of radio sources of far-field region with the flat phase front of electromagnetic wave was shown for bearing angles estimation of radio sources with spherical phase front of electromagnetic wave located in the intermediate-field region relatively to the inputs of linear antenna array of radio direction finder. A modification of Capon's method to deal with the spherical electromagnetic wave front is proposed and some simulation results of bearing angles estimation for several radio sources located simultaneously in the intermediate-field and far-field regions are presented.

**Keywords** Radio source · Spatial spectrum · Electromagnetic wave · Phase front · Sphericity · Linear antenna array · Radio direction finder · Intermediate-field region · Far-field region

## 1 Introduction

In many radio engineering systems (RES), which are related to signal processing by means of linear antenna arrays (LAA) such as radiolocation, hydrolocation, radio intelligence, radio direction finding and wireless communication systems, information about the radio signal direction of arrival (DOA) or its bearing angle from radio source (RS) and number of RS is determined by direction finding methods (otherwise known as correlation direction finding methods) by electromagnetic wave (EMW) phase front radiated from RS [1, 2].

Nowadays, particular attention of scientists is paid to the supperresolution methods of DOA estimation. This methods in their current form, in fact, have received, starting with the works of R. Schmidt which date back to 1979, although

H. Avdieienko (✉) · Y. Yakornov (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Avene 37, Kyiv 03056, Ukraine
e-mail: django2006@ukr.net

Y. Yakornov
e-mail: yakornovits@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_26

their genesis can be traced back to the late 60 s through the works on the estimation of the spectrum in geophysics and the processing of large grids data [3, 4].

Supperresolution methods of DOA estimation are still being developed: new methods are proposed and old improved; the characteristics of the known methods are investigated and compared; there are different variants of modification of the known methods [5].

Correlation direction finding methods can be divided into two classes: by the amount of computational system resources that are used and the time of implementation. The first, "total" class, can include algorithms that require a higher computational cost, which is caused by performing a full two-dimensional search for both coordinates of the azimuth and elevation angles. This includes the following algorithms: BSA scanning algorithm, Capon algorithm, Thermal Noise Algorithm (TNA), generalized Capon method, Multiple Signal Classification (MUSIC) algorithm, and others. The second, "fast" class of algorithms, which require less computational cost, can include algorithms without full dual search. These algorithms typically are efficient modifications of "full" search algorithms. Such algorithms may include: modifications of root-based algorithms, such as Root-MUSIC, the algorithm for evaluating parameters through an rotational invariance technique (ESPRIT), and the Rank Reduction (RARE) algorithm, such as UCA-RARE [5].

## 2 Problem Statement

From a physical point of view, the methods of radio direction finding of RS with angular resolution are based on the same effect as in adaptive antenna array, i.e. coherent compensation (suppression) of interferences signals. The interference signals for each RS are EMW from all others RS that do not coincide with it by DOA. In turn, the term "superresolution" reflects the fact that the methods have a higher angular resolution than the Rayleigh limit

$$\Delta_R = 57.3^\circ \frac{\lambda}{L}, \tag{1}$$

where $\lambda$ is the wavelength of EMW, $L$ is the aperture size of the LAA (Fig. 1a).

That is, they provide direction finding of two RS that fall into the main lobe of the radiation pattern of the M-element LAA even at an angular spacing between this RS less than half the width of the LAA radiation pattern.

Potential characteristics of the direction finding methods are very high for angular coordinates of RS. Thus, the angular resolution in the direction finding between two RS which located in the far-field region [6] at a sufficiently large power of EMW of the latter, can exceed the resolution of the Rayleigh limit (1) in 3… 10 times. However, the potential resolution decreases not only as the signal-to-noise ratio in the LAR channel decreases, but also as the amount of RS increases. The real characteristics of the angular superresolution methods are significantly lower than the potential ones by

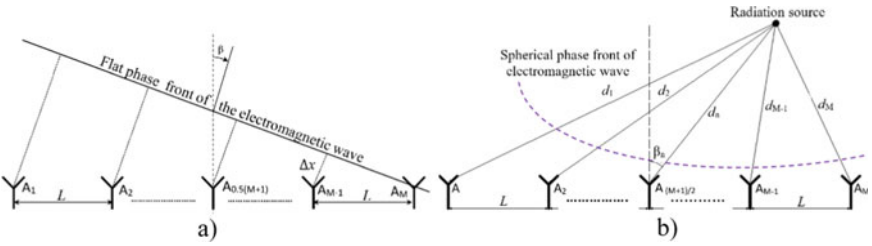



**Fig. 1** Linear antenna array with **a** flat and **b** spherical phase front of EMW

about twice due to the influence of amplitude-phase errors in the receiving channels of the radio direction finder LAA.

The above mentioned set of algorithms for RS DOA estimation, the Capon and MUSIC algorithms are of the greatest interest as they are suitable for use in LAA of arbitrary configuration: one-dimensional, two-dimensional, three-dimensional, regular and irregular etc. In this case, work [7] gives preference to the Capon algorithm, which is simpler in the digital implementation of radio direction finder units and which is almost not inferior in angular resolution to the MUSIC algorithm which popular in practical implementation of radio direction finder.

However, Capon's algorithm for DOA estimation of RS has proven itself well when RS located in the far-field region and its incident EMW has flat wave front at the LAA aperture of radio direction finder (Fig. 1a). But, its DOA estimation effectiveness in determining the bearing angle β of RS located in the intermediate-field region, where there is a incident EMW with spherical phase front at the LAA aperture of radio direction finder (Fig. 1b) has not been studied. The stretch $R_{\text{IFR}}$ of the intermediate-field region can be estimated as [8]

$$R_{NFR} < R_{IFR} < R_{FFR}, \tag{2}$$

where $R_{NFR} \approx 0.6\sqrt{L^3/\lambda}$, $R_{FFR} = 2L^2\cos^2\beta/\lambda$ are boundaries of the near-field and far-field regions respectively.

Methods for determining RS coordinates by the sphericity of the EMW phase front in the intermediate-field region and their practical implementations in radio direction finders are shown in [6], but they work well when there is EMW from single RS received by LAA of radio direction finder.

Therefore, we will consider the possibility of applying the method of angular superresolution according to the Capon algorithm to determine the coordinates (bearing angle and distance) for several RS located both in the far-field and in the intermediate-field regions of LAA of radio direction finder.

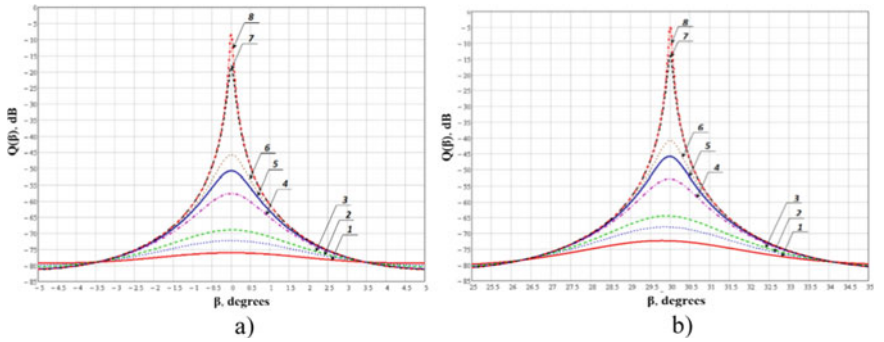



**Fig. 2** Spatial spectrum $Q(\beta)$ for the case of single radio source DOA estimation which located at different distances on the bearing angle: **a** $\beta = 0°$; **b** $\beta = 30°$

## 3 Main Part

The [2] Capon's method is one of the well-known algorithms for direction finding of high resolution by angular coordinates of RS. In most cases, this algorithm involves calculation of the spatial spectrum function (direction finding profile) from the output signals of an M-element LAA (Fig. 2), which excites by EMW with planar phase fronts from RS of the far-field region (Fig. 1a) using the following expression [1]

$$Q(\beta) = \frac{1}{\left| F^2(\beta) \cdot \mathbf{S}_{\alpha}^{\mathbf{H}}(\beta) \cdot \mathbf{R}_{\mathbf{xx}}^{-1} \cdot \mathbf{S}_{\alpha}(\beta) \right|}, \tag{3}$$

where $F(\beta)$ is the beam pattern of a single antenna element of LAR (we will consider further all LAA elements are identical and weakly directed, i.e. $F(\beta) \approx 1$); $\mathbf{R}_{\mathbf{xx}} = E\left[\mathbf{x}(t) \cdot \mathbf{x}^H(t)\right]$– correlation matrix of LAA input signals; $E[]$—averaging operator; $\mathbf{x}(t) = [x_1(t)\ x_2(t) \cdots x_M(t)]^{\mathrm{H}}$—column vector of LAA input signals; $\mathbf{S}_{\alpha}(\beta)$—steering column vector of LAA; H—Hermitian conjugation operation.

When evaluating DOA angles of RS located in far-field region, taking into account the fact that in accordance with Fig. 1a the phase shift $\Delta\varphi(\beta)$ of the harmonic EMW with a flat phase front between adjacent LAA elements will be equal to

$$\Delta\varphi(\beta) = \frac{2\pi\Delta x}{\lambda c} = \frac{2\pi L}{\lambda} \sin\beta \tag{4}$$

and the steering vector column $\mathbf{S}_{\alpha}(\beta)$ can be represented as

$$\mathbf{S}_{\alpha}(\beta) = \left[ \exp(-j(M-1)\Delta\varphi(\beta))\ \exp(-j(M-2)\Delta\varphi(\beta))\ \ldots\ 1 \right]^T. \tag{5}$$

As one can see from expression (3) for the EMW with flat phase front the steering column vector $\mathbf{S}_{\alpha}(\beta)$ functionally depends only on the bearing angle $\beta$ of EMW incidence, and therefore the function of the spatial spectrum $Q(\beta)$ also depends only

on the bearing angle β. Accordingly, by constructing the dependence of Q(β) on DOA bearing angle β, it is possible to obtain information about the number of RS located in the far-field region and their bearing angles. It should be noted that the determination of the RS number and their bearing angles is carried out by the extremum values of the spatial spectrum function $Q(\beta)$.

In order to assess the potential capabilities of the Capon's method, the correlation matrix of input signals at the output of radio direction finder LAA channels under the action of N uncorrelated signals from RS and the uncorrelated white noise presence in each LAA channel based on work [8] can be presented as follows

$$\mathbf{R}_{\mathbf{xx}} = \left( \sum_{n=1}^{N} h_n \mathbf{S}_{\alpha.n} \mathbf{S}_{\alpha.n}^{H} + \mathbf{I} \right) \sigma^2 \tag{6}$$

where $h_n$ is the ratio of the received signal power the from the $n$th RS to the internal white Gaussian noise power in the same LAA channel; $\sigma^2$ is the power of the internal white Gaussian noise in each LAA channel; **I**—identity matrix of $M \times \underline{M}$ dimension; $\mathbf{S}_{\alpha.n}$—column vector of the spatial structure of the EMW radiated from $n$th RS at the LAA linear aperture; $n = 1 \ldots N$, $N$—number of RS.

The column vector $\mathbf{S}_{\alpha.n}$ of the EMW spatial structure in the case of receiving the $n$th signal from the RS located on the bearing angle $\beta_n$ in far-field region of the radio direction finder LAA is determined in accordance with expression (5). In the case of receiving the $n$th signal from the RS, located on the bearing angle $\beta_n$ and distance $d_n$ (Fig. 1b) in the intermediate-field region, column vector $\mathbf{S}_{\alpha.n}(\beta_{\mathrm{n}}, d_{\mathrm{n}})$ is determined according to the expression

$$\mathbf{S}_{\alpha.n}(\beta_n, d_n) = \left[ \exp(-j\Delta\varphi_1(\beta_n, d_n)) \ \exp(-j\Delta\varphi_2(\beta_n, d_n)) \ \ldots \ \exp(-j\Delta\varphi(\beta_n, d_n)) \right]^T, \tag{7}$$

where $\Delta\varphi_m(\beta_n, d_n)$ is the phase shift between the $m$th and the central elements of the LAA (in the case of an odd number of LAR elements) or the phase shift between the $m$th LAA element and the center of the LAA aperture (in the case of an even number of LAA elements), $m = 1 \ldots M$.

The phase shift $\Delta\varphi_m(\beta_n, d_n)$ is determined according to the expression

$$\Delta\varphi_m(\beta_n, d_n) = \frac{2\pi}{\lambda}(d_m(\beta_n, d_n) - d_n), \tag{8}$$

where $d_m(\beta_n, d_n)$ is the distance from the RS with coordinates $(\beta_n, d_n)$ to the $m$th antenna element of the radio direction finder LAA, $m = 1 \ldots M$.

The distance $d_m(\beta_n, d_n)$ for both even and odd number of $M$ elements of the LAR is determined according to the following equation

$$d_m(\beta_n, d_n) = \sqrt{0,25(M + 1 - 2m)^2 L^2 + d_n^2 + (M + 1 - 2m) L d_n \sin\beta_n}, \tag{9}$$

Practical implementation of Capon's method in the radio direction finder digital signal processor in accordance to (3) usually complicated by calculation of true correlation matrix $\mathbf{R}_{xx}$ due to the impossibility of its averaging over an infinite number of implementations or (for ergodic processes) its averaging over an infinite time interval. So, instead of true $\mathbf{R}_{xx}$ matrix its estimation used [2], which can be obtained by temporal sampling of vectors $\mathbf{x}(t)$ at times $t_1 \ldots t_k$ according to the following expression

$$\mathbf{R}_{xx} = \frac{1}{2K} \sum_{k=1}^{K} \mathbf{x}(t_k)\mathbf{x}^H(t_k), \tag{10}$$

where $K$ is the number of time samples in the observation interval.

To analyze the classical Capon's method effectiveness in accordance with expression (3) for the case when RS located in the intermediate-field region of radio direction finder LAA, the results of spatial spectrum function $Q(\beta)$ mathematical simulation in MathCad 14 environment are presented in Fig. 2. The next simulation parameters used: carrier frequency f = 1000 MHz ($\lambda = 0.3$ m); single RS with bearing angles $\beta = 0°$ (Fig. 2a); 30° (Fig. 2b); 75° (Fig. 3a); 80° (Fig. 3b), which is located at different distances from the LAA. This distances are respectively equal to $d_1 = 0.15R_{FFR}$ (curve 1); $d_2 = 0.2R_{FFR}$ (curve 2); $d_3 = 0.25\,R_{FFR}$ (curve 3); $d_4 = 0.5R_{FFR}$ (curve 4); $d_5 = 0.75\,R_{FFR}$ (curve 5); $d_6 = R_{FFR}$ (curve 6); $d_7 = 5R_{FFR}$ (curve 7); $d_8 = 10\,R_{FFR}$ (curve 8). The signal-to-noise ratio in the LAA channels is assumed to be equal to $h = 30$ dB, the bandwidth and noise coefficient of the receiving path of radio direction finder are respectively equal to $\Delta f = 30$ MHz and $NF = 4$ dB, which corresponds to the internal noise power of $\sigma^2 = 1.7 \cdot 10^{-13}$ W (or -97.7 dBm). Equidistant LAA consists of $M = 20$ non-directional elements with distance between elements is $L = 0.5\,\lambda = 0.15$ m. The total size of the LAA aperture is Lp = 2.85 m, and the boundaries of the near-field and far-field regions, respectively, are equal to $R_{NFR} = 5.4$ m; $R_{FFR} = 54$ m.

Analysis of graphs in Figs. 2 and 3 shows that when the RS located in the intermediate-field region at distances of $d = 0, 15R_{FFR} \ldots 0, 25R_{FFR}$, the extremum of the spatial spectrum Q(β), firstly, has a significant angular length, and secondly, has rather small amplitude value of −75… −57 dB in comparison with the case of the same RS location in far-field regions at distances of $d = 5R_{FFR} \ldots 10R_{FFR}$, where the extremum is much narrower and its amplitude is much larger and varies within from −20 to −5 dB. This is especially noticeable at the RS bearing angles of $\beta = -30° \ldots 30°$ (Fig. 2b) at the same signal-to-noise ratio $h = 30$ dB, regardless of the distance of the RS location.

The graphs at Fig. 4 in accordance with expression (3) show the spatial spectrum Q(β) for the case of the two RS direction finding, which located on bearing angles $\beta_1 = -5°$, $\beta_2 = 5°$ (Fig. 4a); $\beta_1 = -3°$, $\beta_2 = 3°$ (Fig. 4b); $\beta_1 = -2°$, $\beta_2 = 2°$ (Fig. 5a); $\beta_1 = -1°$, $\beta_2 = 1°$ (Fig. 5b). In this case, the distance to the 2nd RS was chosen fixed and equal to $d = 10R_{FFR}$ (far-field region), and the distance to the 1st RS in intermediate-field and far-field region was varied and selected the same values as in

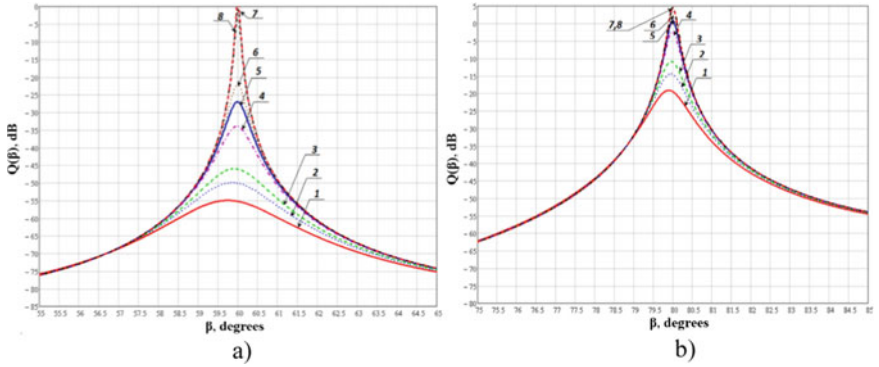



**Fig. 3** Spatial spectrum $Q(\beta)$ for the case of single radio source DOA estimation which located at different distances on the bearing angle: **a** $\beta = 60°$; **b** $\beta = 80°$

Figs. 2 and 3. The signal-to-noise ratio in the LAA channels for the signals from each of the RS for ease of analysis is taken equal to $h_1 = h_2 = 30$ dB.

Analysis of the spatial spectrum in Figs. 4 and 5 shows that when the angular spacing between RS decreases, the angular resolution of the Capon's method is worse the smaller the distance from the second RS to the LAA of the radio direction finder. For example, according to Fig. 4. the DOA of RS which located on the bearing $\beta_1 = -2°$ is not localized for distances equal to $d_1 = 0.15R_{FFR}$; $d_2 = 0.2R_{FFR}$; $d_3 = 0.25\ R_{FFR}$. When the RS is located on the bearing angles $\beta_1$ = -1° and $\beta_2$ = 1° (Fig. 5b), the RS located on the bearing angle $\beta_1 = -1°$ is not localized for the distances to the RS which are equal to $d_1 = 0.15R_{FFR}$; $d_2 = 0.2R_{FFR}$; $d_3 = 0.25R_{FFR}$; $d_4 = 0.5\ R_{FFR}$; $d_5 = 0.75R_{FFR}$; $d_6 = R_{FFR}$.

Similarly to Figs. 4 and 5, the graphs in Fig. 6 shows the spatial spectrum $Q(\beta)$ for the case of three RS direction finding which located on the bearing angles $\beta_1 = -5°$, $\beta_2 = 0°$, $\beta_3 = 5°$ (Fig. 6a); $\beta_1 = -3°$, $\beta_2 = 0°$, $\beta_3 = 3°$ (Fig. 6b); $\beta_1 = -2°$,

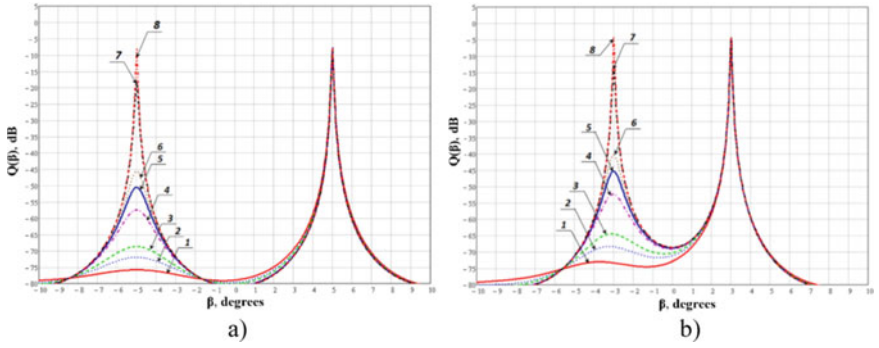



**Fig. 4** Spatial spectrum $Q(\beta)$ for the case of two radio sources DOA estimation which located on the bearing angles: **a** $\beta_1 = -5°$, $\beta_2 = 5°$; **b** $\beta_1 = -3°$, $\beta_2 = 3°$

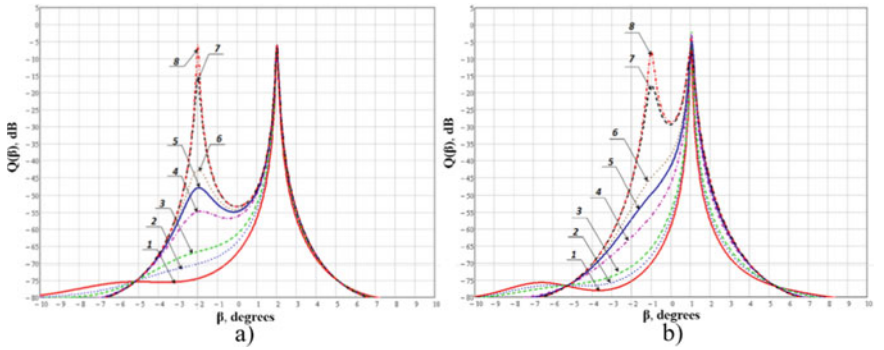



**Fig. 5** Spatial spectrum $Q(\beta)$ for the case of two radio sources DOA estimation which located on the bearing angles: **a** $\beta_1 = -2°$, $\beta_2 = 2°$; **b** $\beta_1 = -1°$, $\beta_2 = 1°$

$\beta_2 = 0°$, $\beta_3 = 2°$ (Fig. 7a); $\beta_1 = -1°$, $\beta_2 = 0°$, $\beta_3 = 1°$ (Fig. 7b). The second RS is located at a fixed distance $d = 10R_{FFR}$ (far-field region), and the distances to the first and third RS are the same and equal to the distances for graphs in Fig. 2. The signal-to-noise ratio in the LAF channels for signals from each RS is still accepted $h_1 = h_2 = h_3 = 30$ dB.

Analysis of the spatial spectrum in Figs. 6 and 7 shows the same regular dependence that for the spatial spectrum in Figs. 4 and 5: the DOA estimation of the RS deteriorates with decreasing angular spacing between the RS and with decreasing distance from the RS to the LAA of the radio direction finder.

Therefore, it can be concluded that in the case when the RS located in the intermediate-field region, the incident EMW phase front from it at the aperture of LAA will be spherical, the spatial spectrum function $Q(\beta)$ analysis shows the RS DOA angle deterioration due to decreasing spatial spectrum function $Q(\beta)$ values in

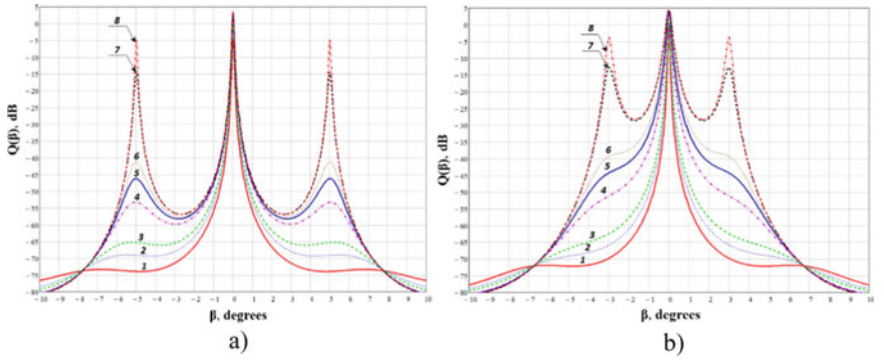



**Fig. 6** Spatial spectrum $Q(\beta)$ for the case of three radio sources DOA estimation which located on the bearing angles: **a** $\beta_1 = -5°$, $\beta_2 = 0°$, $\beta_3 = 5°$; **b** $\beta_1 = -3°$, $\beta_2 = 0°$, $\beta_3 = 3°$

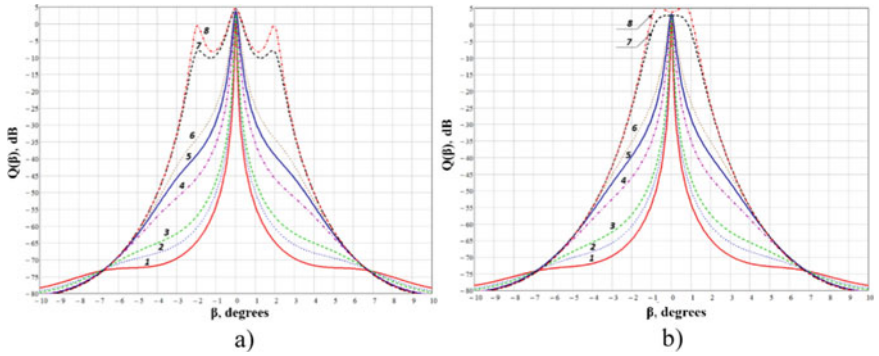



**Fig. 7** Spatial spectrum Q(β) for the case of three radio sources DOA estimation which located on the bearing angles: **a** $\beta_1 = -2°$, $\beta_2 = 0°$, $\beta_3 = 2°$; **b** $\beta_1 = -1°$, $\beta_2 = 0°$, $\beta_3 = 1°$

comparison with the same function for DOA estimation of RS located in far-field region of LAA of radio direction finder.

The difficulty of bearing angle estimation of RS located in the near-field or intermediate-field region of the radio direction finder LAA can be explained by the LAA beam pattern main lobe "blurring", since the LAA is focused for receive the signal from the far-field region using the steering column vector $\mathbf{S}_\alpha(\beta)$, i.e. LAA is defocused relative to the RS which located in the intermediate-field region. Therefore, in order to estimate DOA of RS located in the near-field or intermediate-field region at a distance $d_0$ with Capon's method application it is necessary to focus the LAA on the distance $d_0$ in this region. This, in turn, will lead to defocusing of the LAA relative to the far-field region and worsen the DOA estimation for the RS located in the far-field region.

Therefore, it is necessary to improve (modify) the Capon's method in that way that it can determine the bearing angle of RS with the required accuracy for both RS of near-field and intermediate-field regions for the case of its location at any distance from LAA, even at $d < 0.25R_{FFR}$. To implement this, we propose to modify the mathematical expression (7) of LAA steering column vector $\mathbf{S}_\alpha(\beta)$ in that way, which will take into account the sphericity of the EMW phase front of the expected signal from the RS located in the intermediate-field region at the bearing angle β and range $d$ from LAA. Such modified steering column vector $\mathbf{S}_\alpha(\beta,d)$ is expressed as

$$\mathbf{S}_\alpha(\beta,d) = \left[\exp(-j\Delta\varphi_1(\beta,d))\ \exp(-j\Delta\varphi_2(\beta,d))\ \ldots\ \exp(-j\Delta\varphi(\beta,d))\right]^T. \tag{11}$$

Then, the modified Capon's algorithm will provide for the spatial spectrum function calculation using LAA output signals, which depends on both the bearing angle of the RS and its distance to LAA, i.e. $Q_{IFR}(\beta)$

$$Q_{IFR}(\beta, d) = \frac{1}{\mathbf{S}_{\alpha}^{\mathbf{H}}(\beta, d) \cdot \mathbf{R}_{\mathbf{xx}}^{-1} \cdot \mathbf{S}_{\alpha}(\beta, d)}. \quad (12)$$

Therefore, after mathematical expression of the LAA steering vector $\mathbf{S}_{\alpha}(\beta,d)$ determination, which allows to take into account the EMW phase front sphericity from the RS for a given LAA geometric configuration, the algorithm for bearing angle $\beta_n$ and the distance $d_n$ estimation for the $n$th RS located in the intermediate-field region of LAA can be represented in Fig. 8 [9, 10]. This algorithm includes the principle of distance range $\Delta d_{(k-1)k}$ dividing into $L$ of uniform subbands with a step $\delta d$ ($L = \Delta d/\delta d$) (Fig. 9).

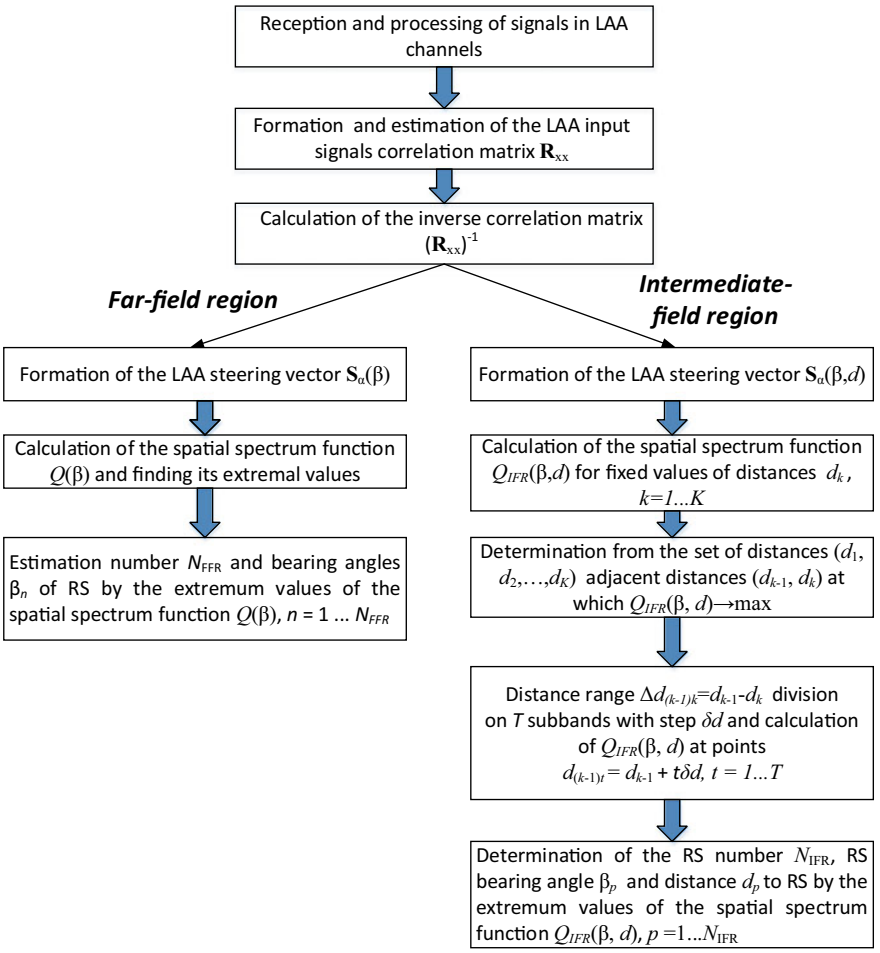



**Fig. 8** Main stages of the proposed modified Capon's method for coordinate determination of the RS located in the intermediate-field region

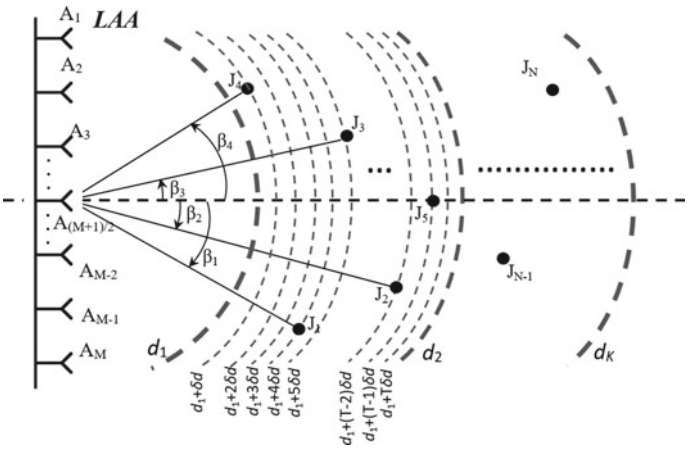



**Fig. 9** Linear antenna array and distance to RS for its focusing

Schematic implementation of radio direction finder to determine the coordinates of several RS by the classical and advanced Capon's method consists of cascaded: LAA, receiver unit, analog-to-digital converter and microprocessor device (Figs. 10 and 11). The latter uses an algorithm, the block diagram of which is shown in Fig. 12. Consider some simulation results using this algorithm.

The graphs in Fig. 13 presents the spatial spectrum $Q_{IFR}(\beta, d)$,which calculated and plotted in accordance with (12) for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and distance $d_0 = 0{,}15R_{FFR}$ with sequential LAA focusing with a step $\delta d = 0{,}01R_{FFR}$ using the steering vector $\mathbf{S}_\alpha(\beta, d)$ at the distance of $d_1 = 0.1R_{FFR}$ (curve 1); $d_2 = 0.11R_{FFR}$ (curve 2); $d_3 = 0.12R_{FFR}$ (curve 3); $d_4 = 0.13R_{FFR}$ (curve 4); $d_5 = 0.14R_{FFR}$ (curve 5); $d_6 = 0.15R_{FFR}$ (curve 6) (Fig. 13a) and $d_6 = 0.15R_{FFR}$ (curve 6); $d_7 = 0.16R_{FFR}$ (curve 5); $d_8 = 0.17R_{FFR}$ (curve 4); $d_9 = 0.18\,R_{FFR}$ (curve 3); $d_{10} = 0.19R_{FFR}$ (curve 2); $d_{11} = 0.2\,R_{FFR}$ (curve 1) (Fig. 13b).

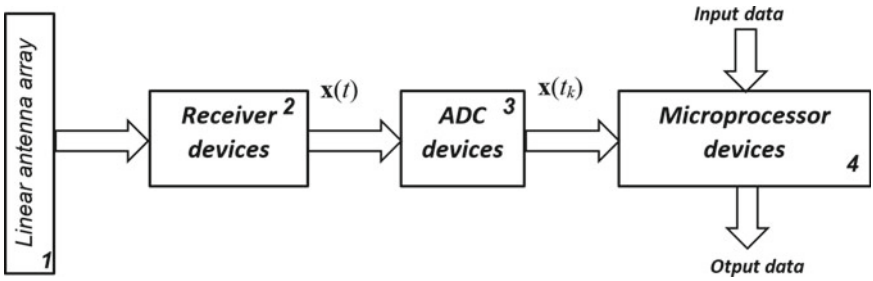



**Fig. 10** Generalized block diagram of the radio directing finder for several RS coordinate determination with usage of the modified Capon's method

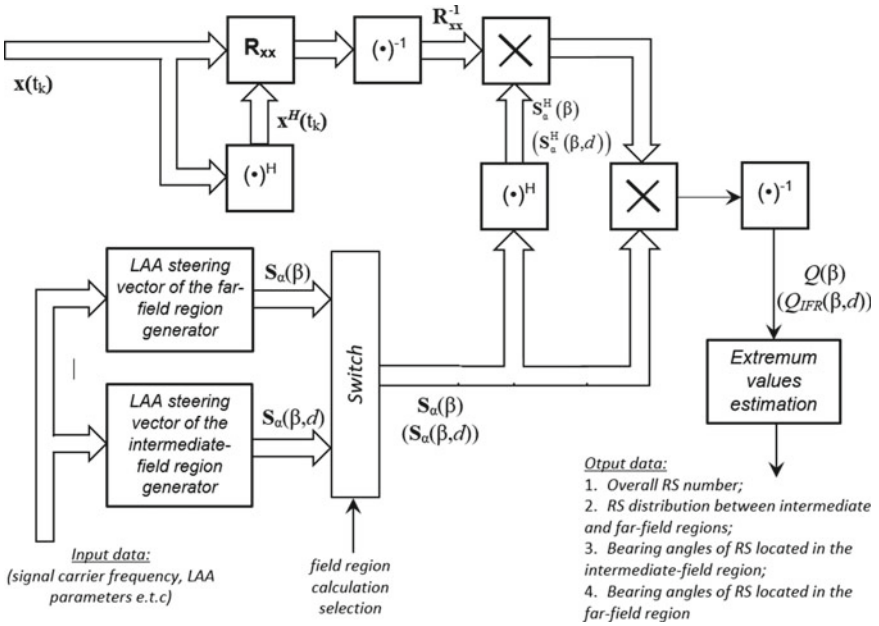



**Fig. 11** Simplified block diagram of the microprocessor unit of the radio direction finder for the modified Capon's method implementation

Similarly to Fig. 13 graphs in Fig. 14 presents the spatial spectrum $Q_{IFR}(\beta, d)$, which calculated for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and distance $d_0 = 0.2R_{FFR}$ with sequential LAA focusing with a step $\delta d = 0.01R_{FFR}$ within $(d_0 - 5\delta d, d_0 + 5\delta d)$ range of distance, i.e. $d_1 = 0.15R_{FFR}$ (curve 1); $d_2 = 0.16R_{FFR}$ (curve 2); $d_3 = 0.17R_{FFR}$ (curve 3); $d_4 = 0.18R_{FFR}$ (curve 4); $d_5 = 0.19R_{FFR}$ (curve 5); $d_6 = 0.2R_{FFR}$ (curve 6) (Fig. 14a) and $d_6 = 0.2R_{FFR}$ (curve 6); $d_7 = 0.21R_{FFR}$ (curve 5); $d_8 = 0.22R_{FFR}$ (curve 4); $d_9 = 0.23\,R_{FFR}$ (curve 3); $d_{10} = 0.24R_{FFR}$ (curve 2); $d_{11} = 0.25\,R_{FFR}$ (curve 1) (Fig. 14b).

Graphs in Fig. 15 presents the spatial spectrum $Q_{IFR}(\beta, d)$, which calculated for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and distance $d_0 = 0{,}5R_{FFR}$ with sequential LAA focusing with a step $\delta d = 0.01R_{FFR}$ within $(d_0 - 5\delta d, d_0 + 5\delta d)$ range of distance, i.e. $d_1 = 0.45R_{FFR}$ (curve 1); $d_2 = 0.46R_{FFR}$ (curve 2); $d_3 = 0.47R_{FFR}$ (curve 3); $d_4 = 0.48R_{FFR}$ (curve 4); $d_5 = 0.49R_{FFR}$ (curve 5); $d_6 = 0.5R_{FFR}$ (curve 6) (Fig. 15a) and $d_6 = 0.5R_{FFR}$ (curve 6); $d_7 = 0.51R_{FFR}$ (curve 5); $d_8 = 0.52R_{FFR}$ (curve 4); $d_9 = 0.53\,R_{FFR}$ (curve 3); $d_{10} = 0.54R_{FFR}$ (curve 2); $d_{11} = 0.55R_{FFR}$ (curve 1) (Fig. 15b).

The analysis of the spatial spectrum presented in Figs. 13, 14 and 15, shows that the extremum value of the spatial spectrum function $Q_{\text{IRF}}(\beta, d)$ and accordingly the minimum of its width will be on the RS bearing angle only for the case of coincidence of the distance to the RS ($d_0 = 0.15R_{FFR}; 0.2R_{FFR}; 0.5\,R_{FFR}$) with the range on which the LAA of radio direction finder focuses ($d_6 = 0.15R_{FFR}; 0.2R_{FFR}; 0.5\,R_{FFR}$). The

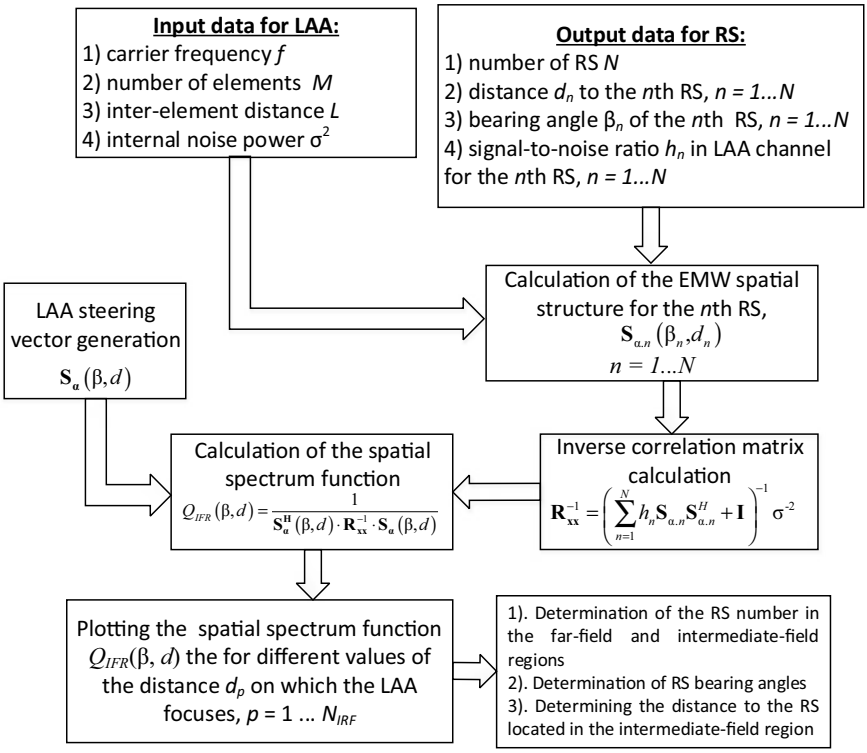



**Fig. 12** The mathematical simulation block diagram for investigation potential characteristics of the modified Capon's method

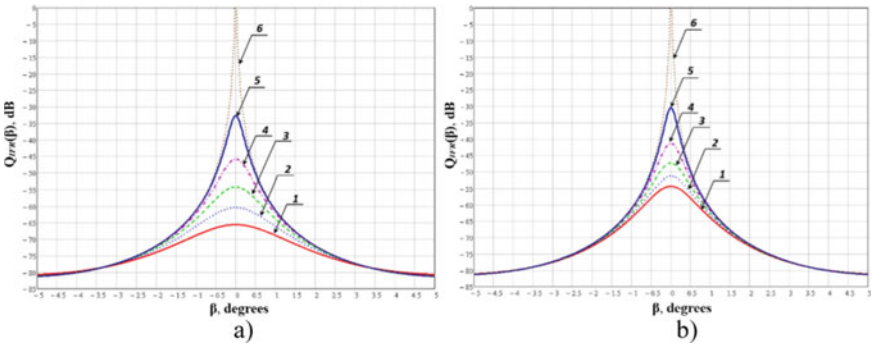



**Fig. 13** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and distance $d_0 = 0.15R_{FFR}$ with sequential LAA focusing with a step $\delta d = 0.01R_{FFR}$

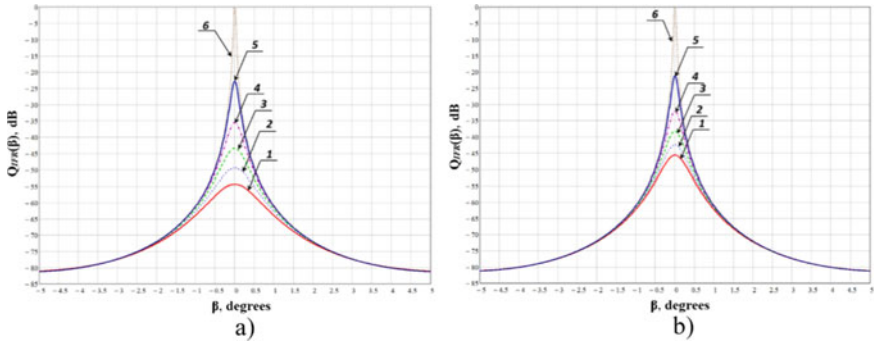



**Fig. 14** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and distance $d_0 = 0.2R_{FFR}$ with sequential LAA focusing with a step $\delta d = 0.01R_{FFR}$

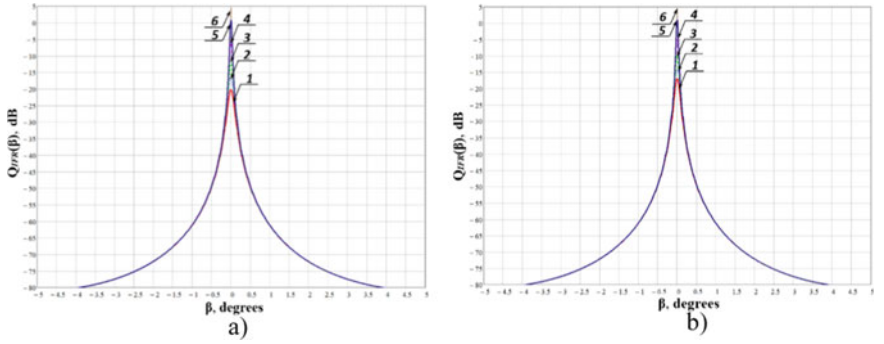



**Fig. 15** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and distance $d_0 = 0.5R_{FFR}$ with sequential LAA focusing with a step $\delta d = 0.01R_{FFR}$

same conclusion can be reached if we analyze spectrum function $Q_{\text{IRF}}(\beta, d)$ for other angular directions.

The graphs in Figs. 16 and 17 presents the spatial spectrum $Q_{IFR}(\beta, d)$ calculated and plotted according to (12) for the case when RS bearing angle is $\beta_0 = 0°$ depending on the distance $d = d_0$ to the RS. The RS distance is equal to $d_0 = 0.15R_{FFR}$ (curve 1), $d_0 = 0.25\ R_{FFR}$ (curve 2), $d_0 = 0.5R_{FFR}$ (curve 3), $d_0 = 0.75\ R_{FFR}$ (curve 4), $d_0 = R_{FFR}$, (curve 5), $d_0 = 5R_{FFR}$ (curve 6), $d_0 = 10R_{FFR}$ (curve 7) and signal-to-noise ratio in the LAA channel equal to: $h_1 = 30$ dB (Fig. 16a); $h_1 = 20$ dB (Fig. 16b); $h_1 = 10$ dB (Fig. 17a); $h_1 = 0$ dB (Fig. 17b).

Analysis of the simulation results presented in Figs. 16 and 17 shows that at the RS distance $d_0$ in the intermediate-field region the spatial spectrum function $Q_{\text{IRF}}(\beta, d)$ has a extremum value. As the RS distance $d_0$ approaches to the far-field region, the distance resolution of the RS worsen because the EMW phase front sphericity

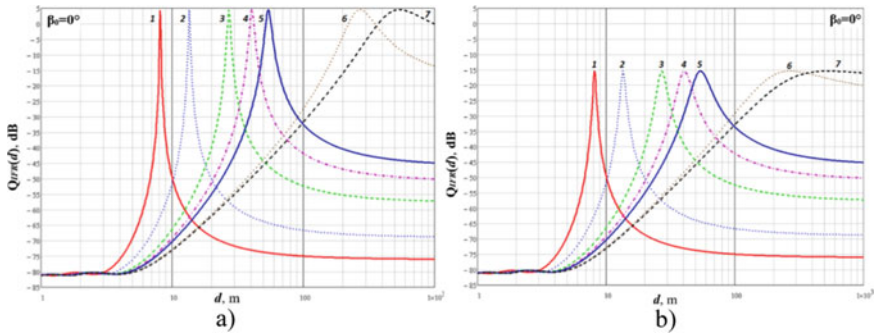



**Fig. 16** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and various distance $d = d_0$: **a** $h_1 = 30$ dB; **b** $h_1 = 20$ dB

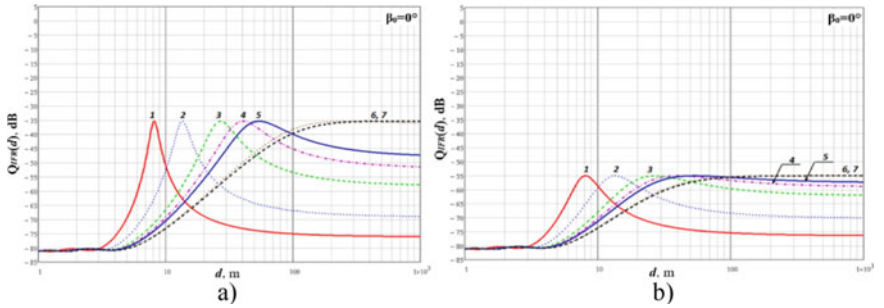



**Fig. 17** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of single RS location in the intermediate-field region on the bearing angle $\beta_0 = 0°$ and various distance $d = d_0$: **a** $h_1 = 10$ dB; **b** $h_1 = 0$ dB

at the LAA aperture decreases, which leads to an expansion of the spatial spectrum function $Q_{\text{IRF}}(\beta, d)$ extremum. In addition, the distance resolution depends on the value of the parameter $h_1$ (signal-to-noise ratio). At large values of the parameter $h_1$ ($\geq$ 30 dB) it is possible to roughly estimate the distance even for those RS which located outside the far-field region boundary, i.e. for a RS distance equal to $d_0 = 5\ldots10R_{FFR}$. As the parameter $h_1$ decreases, the extremum expands and, accordingly, the estimation of the RS distance worsen even for those RS which located in the intermediate-field region.

The graphs in Fig. 11 shows the spatial spectrum calculated according to (12) for two RS located in different wave regions on bearing angles $\beta_1 = -5°$ (intermediate-field region) and $\beta_2 = 5°$ (far-field region) for the case of the LAA focusing in far-field region (curve 2) and LAA focusing in the intermediate-field region (curve 1) at a distance: $d_0 = 0.5R_{FFR}$; (Fig. 18a); $d_0 = R_{FFR}$ (Fig. 18b). In Fig. 19 are all the same, but with the location of the two RS on bearing angles $\beta_1 = -1°$ and $\beta_2 = 1°$. As can be seen from Figs. 18 and 19 in both cases, DOA of both RS are reliably determined taking into account that the Rayleigh boundary equals to

$$\Delta_R = 57,3^\circ \frac{\lambda}{L} = 57,3^\circ \frac{\lambda}{20 \cdot 0,5\lambda} = 5,73^\circ. \quad (13)$$

Comparison of Figs. 18 and 19 with Figs. 4 and 5 shows that the proposed modified Capon's method in contrast to the classical Capon's method allows to determine not only RS bearing angle, but also the distance to RS, which is located in the intermediate-field region of the LAA. This is due to the possibility of changing the phase distribution (EMW spatial structure or EMW phase front shape) of the expected signal given by the steering vector $\mathbf{S}_\alpha(\beta, d)$.

It is interesting to study the spatial spectrum for cases where two RS are located on the same bearing angle $\beta_1 = \beta_2 = 0^\circ$, but at different distances from the radio direction finder, for example: 1) both RS are placed in the intermediate-field region; 2) the first RS located in the intermediate-field region, and the other RS located in far-field region.

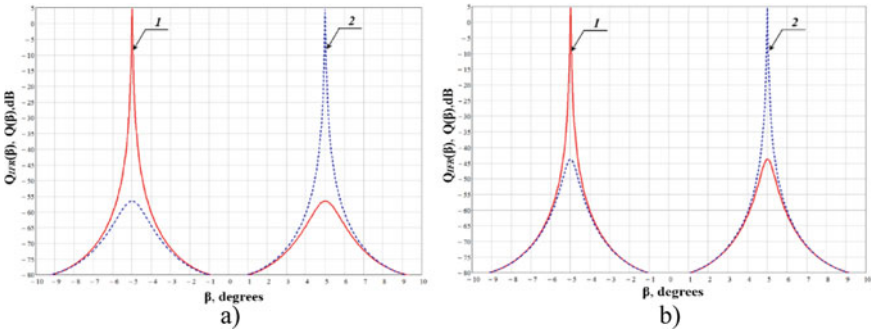



**Fig. 18** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of two RS location in different wave regions on the bearing angles $\beta_1 = -5^\circ$ (intermediate-field region) and $\beta_2 = 5^\circ$ (far-field region)

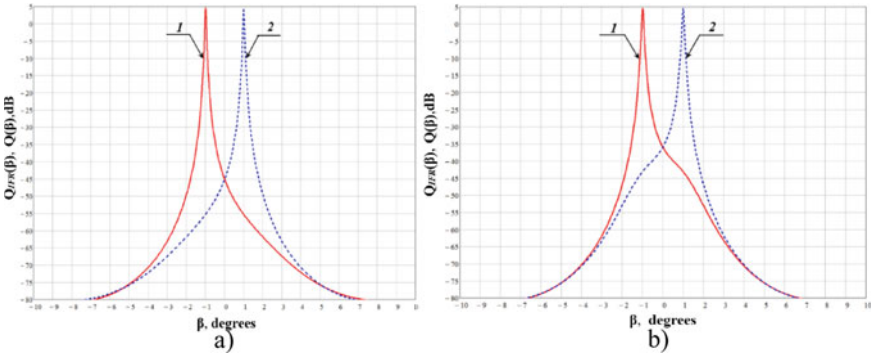



**Fig. 19** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of two RS location in different wave regions on the bearing angles $\beta_1 = -1^\circ$ (intermediate-field region) and $\beta_2 = 1^\circ$ (far-field region)

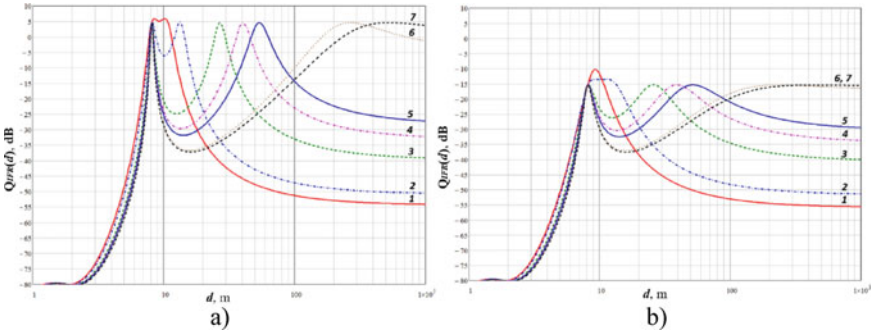



**Fig. 20** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of two RS location in the interme-diate-field region on the bearing angle $\beta = 0°$ and various distances $d$: **a** $h_1 = 30$ dB; **b** $h_1 = 20$ dB

The graphs in Fig. 20 presents the spatial spectrum function $Q_{IFR}(\beta, d)$ for two RS placed on one bearing $\beta_1 = \beta_2 = 0°$, while the distance to the 1st RS was constant and equals to $d_1 = 0.15R_{FFR}$ and the distance to the 2nd RS was equal to: $d_2 = 0.2R_{FFR}$ (curve 1); $d_2 = 0.25R_{FFR}$ (curve 2); $d_2 = 0.5R_{FFR}$ (curve 3); $d_2 = 0.75\ R_{FFR}$ (curve 4); $d_2 = R_{FFR}$ (curve 5); $d_2 = 5\ R_{FFR}$ (curve 6); $d_2 = 10\ R_{FFR}$ (curve 7). The value signal-to-noise parameter for both RS was chosen equal to $h_1 = h_2 = 30$ dB (Fig. 20a) and h1 = h2 = 20 dB (Fig. 20b).

Analysis of the curves in Fig. 20 shows the possibility of the distance estimating to each of the two RS located in the intermediate-field region. It is possible, firstly, when there are a high value of the parameters $h_1$, $h_2$, and secondly, for such mutual distance diversity of RS relative to each other, which will be sufficient for the EMW phase front sphericity from these RS differs significantly.

Figure 21 shows the spatial spectrum for DOA estimation of three RS located in different field regions on the bearing angles $\beta_1 = -5°$ (intermediate-field region), $\beta_2 = 0°$ (far-field region) and $\beta_3 = 5°$ (intermediate-field region) when focusing the LAA to the far-field region (curve 2) and LAA focusing to the distances $d_1 = 0.15R_{FFR}$ and $d_1 = 0.25R_{FFR}$ of the intermediate-field region (curve 1). In Fig. 22 are all the same as in Fig. 21, but with the location of the three RS on bearing angles $\beta_1 = -1°$, $\beta_2 = 0°$ and $\beta_3 = 1°$.

In the case of three RS located in different wave field zones, as shown in Figs. 20 and 21 in comparison with Figs. 6 and 7 angular separation of the RS by DOA and its distance determination are possible even at their diversity on a bearing angle each other that equals to $\Delta\beta = \pm 1°$. But such separation is available on condition that distance to LAA does not exceed $d_0 = 0.25R_{FFR}$ (Fig. 22b). At higher values of the RS distance in the intermediate-field region (RS moves in the direction of the far-field region boundary) with such a small angular spacing, the peaks of the spatial spectrum function "blur" and the determination of bearing angles and distances to each RS of the intermediate-field region becomes impossible. At the same time, RS located in the far-field region is reliably localized in all these cases.

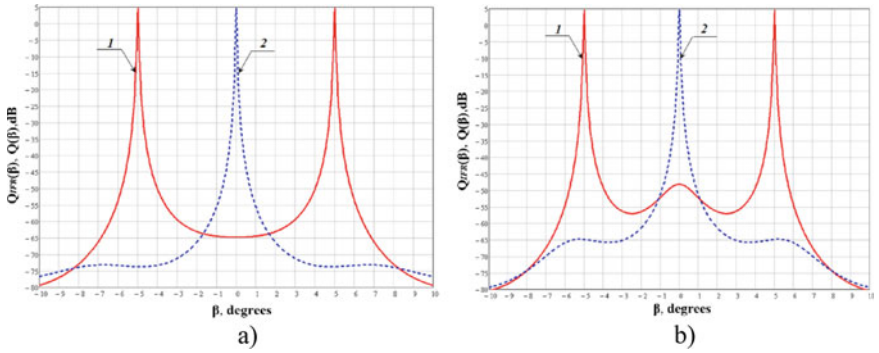



**Fig. 21** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of three RS location in different wave regions on the bearing angles $\beta_1 = -5°$ (intermediate-field region), $\beta_2 = 0°$ (far-field region) and $\beta_3 = 5°$ (intermediate-field region)

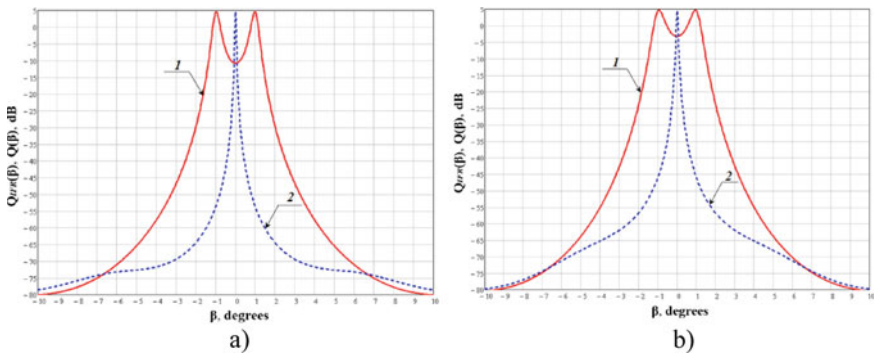



**Fig. 22** Spatial spectrum $Q_{IFR}(\beta, d)$ for the case of three RS location in different wave regions on the bearing angles $\beta_1 = -1°$ (intermediate-field region), $\beta_2 = 0°$ (far-field region) and $\beta_3 = 1°$ (intermediate-field region)

Theoretical studies presented in article [5] show that it is possible to ensure a further increase in the angular coordinate (bearing angle) resolution by using the generalized Capon's method. Then, for RS located in the intermediate-field region we can propose a generalized improved Capon's method, the mathematical representation of the spatial spectrum function $Q_{M.IFR}(\beta, d)$ of which based on expression (10) and will express like

$$Q_{M.IFR}(\beta, d) = \frac{1}{\mathbf{S}_{\boldsymbol{\alpha}}^{\mathbf{H}}(\beta, d) \cdot \mathbf{R}_{\mathbf{xx}}^{-N} \cdot \mathbf{S}_{\boldsymbol{\alpha}}(\beta, d)} \tag{14}$$

where $\mathbf{R}_{\mathbf{xx}}^{-N} = \mathbf{R}_{\mathbf{xx}}^{-1} \cdot \mathbf{R}_{\mathbf{xx}}^{-1} \cdot .... \cdot \mathbf{R}_{\mathbf{xx}}^{-1}$ is the consequent multiplication of $N$ inverse correlation matrices.

Then, for the case of the DOA estimation of three RS (two of them located in the intermediate-field region on bearing angles $\beta_1 = -1°$, $\beta_3 = 1°$ and distance $d_0$, the third RS located in the far-field region on bearing angle $\beta_2 = 0°$) and using expression (14) for N = 1 (curve 1), N = 2 (curve 2) and N = 3 (curve 3), taking into account that signal-to-noise ratio in receiver channels of LAA equals to $h_1 = h_2 = h_3 =$ 30 dB we obtain the spatial spectrum function $Q_{M.IFR}(\beta, d)$ for the intermediate-field and far-field regions, which are shown in Fig. 23 ($d_0 = 0.5R_{FFR}$), and Fig. 24 ($d_0 = R_{FFR}$).

A comparison of the graphs obtained for the case of applying the improved Capon's method (Fig. 23a, curve 1) with the spatial spectra obtained for the case of applying the generalized improved Capon's method (Fig. 23a, curve 2 and 3) shows that the latter method really allows to clearly distinguish angular position of two RS located in the intermediate-field region at distance $d_0 = 0.5R_{FFR}$. The same

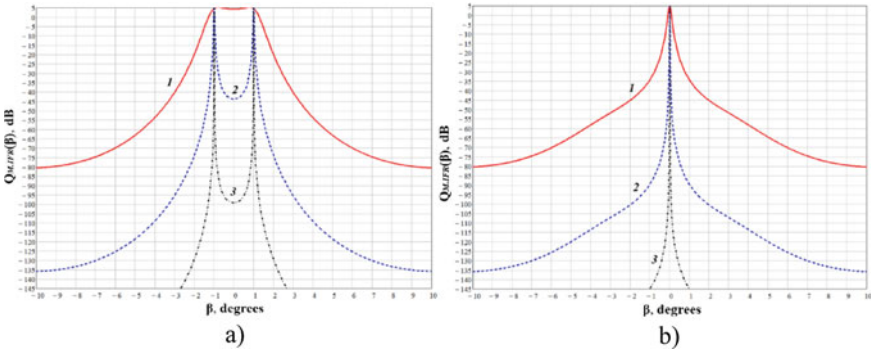



**Fig. 23** Spatial spectrum $Q_{M.IFR}(\beta, d)$ for the case of three RS location in different wave regions on the bearing angles **a** $\beta_1 = -1°$, $\beta_3 = 1°$ (intermediate-field region, distance to two RS is equals to $d_0 = 0.5R_{FFR}$); **b** $\beta_2 = 0°$ (far-field region)

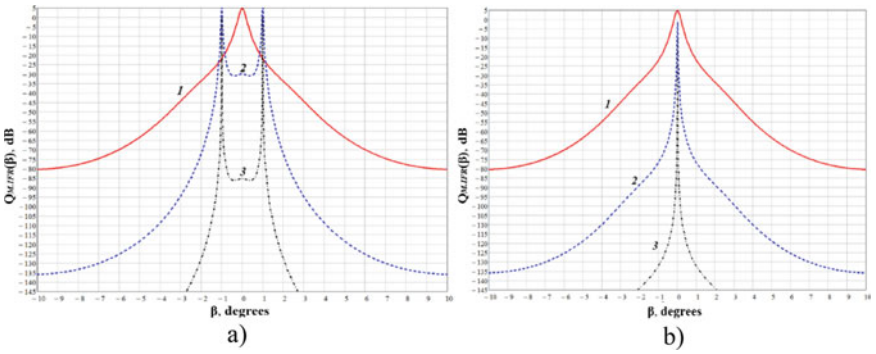



**Fig. 24** Spatial spectrum $Q_{M.IFR}(\beta, d)$ for the case of three RS location in different wave regions on the bearing angles **a** $\beta_1 = -1°$, $\beta_3 = 1°$ (intermediate-field region, distance to two RS is equals to $d_0 = R_{FFR}$); **b** $\beta_2 = 0°$ (far-field region)

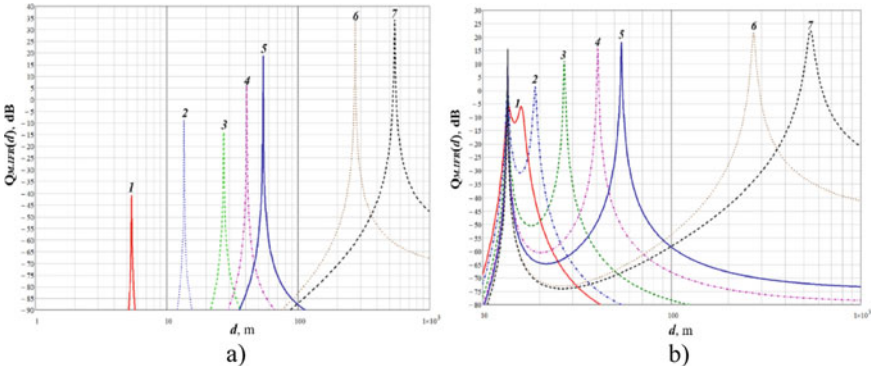



**Fig. 25** Spatial spectrum $Q_{M.IFR}(\beta, d)$ for the case of simple RS location **a** and two RS location in the intermediate-field and far-field region **b** at various distances $d$

distinction of the angular position of the two RS located in the intermediate-field region, but with some deterioration due to the expansion of the spatial spectrum function extremums can be provided for the distance equal to $d_0 = R_{FFR}$ (Fig. 24a). At the same time, RS located in the far-field region is reliably localized in all these cases (Fig. 23b, Fig. 24b, curve 2 and 3).

It is important to note that the generalized improved Capon's method can also provide a distance resolution of the RS location, even for the those RS which located in the far-field region. This is confirmed by the results of mathematical modeling of the spatial spectrum function (14) for the case of the simple RS location at distances of $d_0 = 0.2R_{FFR}$ (curve 1); $d_0 = 0.25R_{FFR}$ (curve 2); $d_0 = 0.5R_{FFR}$ (curve 3); $d_0 = 0.75\,R_{FFR}$ (curve 4); $d_0 = R_{FFR}$ (curve 5); $d_0 = 5R_{FFR}$ (curve 6); $d_0 = 10R_{FFR}$ (curve 7), presented in Fig. 25a. In Fig. 25b are all the same as in Fig. 25a, but with the location of the second RS at the constant distance equals to $d_1 = 0.15R_{FFR}$.

A comparison of Fig. 25a with Fig. 16a, as well as Fig. 25b with Fig. 20a shows that generalized improved Capon method's application generate clear extremum of the spatial spectrum function at the RS distance, even for the RS located in far-field regions at distances $d_0 = 5R_{FFR}$ and $d_0 = 10R_{FFR}$. So, there is a possibility of a more accurate estimation of the distance to RS, which is located even at a distance of several units-tens of the far-field region boundary.

## 4 Conclusions

1. The classic Capon's method for direction-finding of the RS located in the intermediate-field region at distances of $d = 0.15R_{FFR} \ldots 0.0.25\,R_{FFR}$ from LAA of radio direction finder allows to obtain a spatial spectrum function Q(β), which extremum, firstly, has a significant angular length, and secondly, has a small

amplitude value −75… −57 dB compared to the case of the DOA determination of the same RS but located in the far-field region.

2. In the case of two RS located in different wave field zones and having a relatively small mutual difference in angular coordinate, the angular resolution of the classical Capon's method is worse the smaller the distance from the RS located in the intermediate-field region to the LAA of radio direction finder. This deterioration in the angular resolution of the Capon's method is explained only by the difference between the phase fronts of EMW from RS of other wave field zones at the LAA aperture, i.e. not taking into account the EMW phase front sphericity from RS located in the intermediate-field region in the mathematical model of the LAA steering column vector $\mathbf{S}_{\alpha}(\beta)$.
3. To ensure the direction-finding of the RS located in the intermediate-field region and far-field regions, it is proposed to improve (modify) the classical Capon method by introducing a mathematical model of the EMW with spherical phase front instead of the model EMW with the flat phase front in the steering column vector $\mathbf{S}_{\alpha}(\beta, d)$ of LAA.
4. The simulation results using the modified Capon's method show the possibility of determining the bearing angles of RS located in the far-field region, as well as bearing angles together with the distances to them for RS located in the intermediate-field region.
5. In the future, the application of the proposed generalized modified Capon's method will significantly improve the resolution of the RS localization at angular coordinates when placing the several RS in the intermediate-field region at the same distance, but with small angular diversity or improving the distance resolution for the case of several RS location with the same bearing angles but different distances to LAA.

## References

1. Capon J (1969) High-resolution frequency-wavenumber spectrum analysis. Proc IEEE 57:1408–1418. https://doi.org/10.1109/PROC.1969.7278
2. Gabriel W (1980) Spectral analysis and adaptive array superresolution techniques. Proc IEEE 68:654–666. https://doi.org/10.1109/PROC.1980.11719
3. Schmidt R (1979) Multiple emitter location and signal parameter estimation. In: Proceedings of RADC spectrum estimation workshop, Rome, NY, pp 243–258
4. Schmidt RO (1986) Multiple emitter location and signal parameter estimation. IEEE Trans Ant Propag 34:276–280. https://doi.org/10.1109/TAP.1986.1143830
5. Korobkov, M.A.: Korrelyatsionnyie metodyi pelengovaniya istochnikov izlucheniya. Molodoy uchyonyiy, 13(72), pp. 55–58 (2014) (in Russian)
6. Avdeyenko GL, Lipchevskaya IL, Yakornov EA (2012) Phase systems of determining coordinates of radiation source with harmonic signal in Fresnel zone. Radioelectron Commun Syst 55:65–74. https://doi.org/10.3103/S0735272712020021
7. Ratyinskiy MV (2003) Adaptatsiya i sverhrazreshenie v antennyih reshetkah. Radio i svyaz, Moskva, p 200. (in Russian)

8. Schesnyak SS, Popov MP (1995) Adaptivnyie antennyi. Sankt-Peterburg: VIKKA im. A.F. Mozhayskogo, p 611
9. Sposib vyznachennia mistseroztashuvannia dzherela radiovyprominiuvannia v blyzhnii zoni(2017) pat. 113916 Ukraina, MPK G01S 5/08, no. 201606780, zaiavl. 09.09.2016, opubl. 27.02.2017, Biul. no. 4 (in Russian)
10. Avdeyenko G, Yakornov E, Korsak V (2016) Spatial processing algorithm of radiation sources in the near and intermediate zones of linear antenna array for monitoring systems. In: 2016 IEEE international scientific conference (UkrMiCo 2016): Conference proceedings, Kyiv, Ukraine, NTUU "KPI", 11–15 September. https://doi.org/10.1109/UkrMiCo.2016.7739628

# Different Approaches for Analytic and Numerical Estimation of Operation Temperature of Cooled Cathode Surface in High Voltage Glow Discharge Electron Guns

**Igor Melnyk, Serhii Tuhai, Mykola Surzhykov, Iryna Shved, Vitaliy Melnyk, and Dmytro Kovalchuk**

**Abstract** Analytical and numerical estimation of operation temperature of the surface of cooled cathode in the high voltage glow discharge electron guns is provided in the chapter. All theoretical presumptions are based both on analytical and numerical solving of heat partial differential equation. The different constructions of the cathode cooling system are also considered. Obtained analytical solutions are based both on transforming of Boltzmann Thermodynamic Equation and on considering Bessel and Integral Exponential functions expansion for nuclear of heat equation. Obtained numerical solutions are based on applying of CAM Solidworks simulation software instrument. Experimental estimations of the cathode energetic efficiency for different metals and sort of residual gases are also given. The simulation results are shown, that with power of electron gun till 500 kW and using aluminum as cathode material, for the suitable construction of cooling system, the operation temperature of the cathode surface not greater than 200 °C. It is also proving, that difference between the preliminary estimation with using Boltzmann Thermodynamic Equation and numerical calculation with using CAM Solidworks instrument is usually

I. Melnyk (✉) · S. Tuhai (✉) · M. Surzhykov (✉) · I. Shved (✉) · V. Melnyk (✉)
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Avenue 37, building 12, Kyiv 03056, Ukraine
e-mail: imelnik@phbme.kpi.ua

S. Tuhai
e-mail: sbtuhai@gmail.com

M. Surzhykov
e-mail: nikolajsurzikov@gmail.com

I. Shved
e-mail: shvettd@gmail.com

V. Melnyk
e-mail: vmelnik@phbme.kpi.ua

D. Kovalchuk (✉)
Joint Stock Company, Scientific and Industrial Association "Chervona Hvylia", vul. Bozhenka, 15, Kyiv 03680, Ukraine
e-mail: dv_kovalchuk@yahoo.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5_27

smaller, than 30%, and simple analytical approach given the correct value of cathode surface temperature.

**Keywords** High voltage glow discharge · Glow discharge electron gun · Electron beam technologies · Cold cathode · Heat equation · Boltzmann thermodynamic equation · Solidworks CAM software

## 1 Introduction

Electron beam technologies are very effective and important today and it's applied in different branches of modern industry. Mostly it caused by such advantages of these technologies, as high total power and power density of electron beam, simplicity of spatial moving of electron beam by using electric and magnetic fields, and providing the technological operation in vacuum to ensure the cleanliness and high quality of obtained products. Development and industrial application of modern electron-beam technologies requires today both improvement of traditional electron sources with heated cathodes [1–4] and elaboration the new types of electron sources, based on other physical principles [5–8].

Today the main branches of industry, where the heated treatment of items by powerful electron beams are widely used, are metallurgy, mechanical engineering, airspace industry, automotive industry, instrument making and electronics [1–4, 9–18]. For example, in the electronic industry heating by the powerful electron beams can be applied in such complex technological processes, as growth of semiconductors nanocrystals [13, 14] and obtaining thin complex ceramics films, used in band high-frequency filters in the modern communication devices for mobile systems and computer networks [15–18].

Among the new types of electron sources, which are at the recent time widely used in industry for refining the refractory metals [2], welding of thin-wall items at the instrument-making industry [7], as well as for deposition of chemically-complex ceramic films in the airspace, automotive and electronic industry [9, 10, 12], the High Voltage Glow Discharge Electron Guns (HVGDEG) [6, 7] have to be pointed out. One of the recent successful applications of HVGDEG is additive manufacture, namely, printing the metallic details with the complex spatial geometry by uninterrupted melting of moving wire [19]. The principle of operation of HVGDEG's is maintaining the gas discharge between cathode and anode at acceleration voltage range of few or tens kV and pressure of residual gas few Pa [7, 8]. The main advantages of HVGDEG relatively to the traditional electron sources with the heated cathodes are follows [7, 8]:

1. Possibility of operation in the soft vacuum in the medium of different gases, including noble and active ones.
2. Relative simplicity of guns' construction, including the possibility of its disassembling for changing the spare details and further repairing. The most common

malfunction for such type of guns is sputtering and contamination of the operation cathode surface.
3. Relative simplicity and cheapness of necessary technological evacuation equipment.
4. High productivity of electron-beam heating technological process. This advantage is especially noticeable in the case of treatment of the items with the linear-like or ring-like geometry by using the profile electron beams, formed in the low-pressure discharge with the suitable spatial geometry of electrodes [7, 19, 20].
5. Simplicity of control of electron beam current both aerodynamically by changing the pressure in discharge chamber [21] and electrically by applying the potential on the additional electrode, located in anode plasma, and, as a result, changing the concentration of ions and electrons in the discharge gap [22, 23].

Among the possible application of HVGDEG in electronics industry refining of silicon [3], growth the nanocrystals for low-temperature devices [13, 14], as well as obtaining of thin ceramic films for the band high-frequency filters in communication devices [15–18], have to be pointed out.

The photographs of elaborated HVGDEG's with different power [7], as well as photograph of elaborated HVGDEG with the electronic system for aerodynamic control of beam current and electromagnetic valve [21] are presented at Fig. 1.

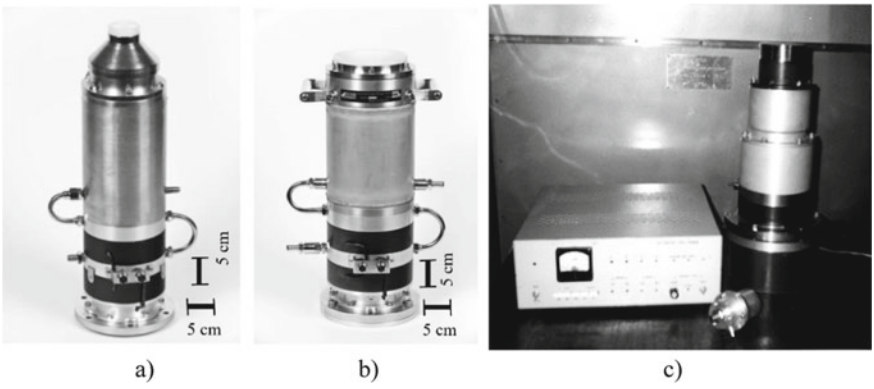



**Fig. 1** Photographs of high voltage glow discharge electron guns' assemblies and electromagnetic valve for automatic control of beam current: **a** the HVGDEG with power of electron beam 100 kW [7], **b** the HVGDEG with power of electron beam 300 kW [7], **c** the HVGDEG with electromagnetic valve and electronic control system for stabilization the current of electron beam, the beam power is 30 kW [21]

## 2 Statement of Considered Problem and Motivation

It is generally clear, that for effective elaboration of new improved constructions of high voltage glow discharge (HVGD) electron guns and its further applying in the modern electron-beam technological equipment, improving the stability and reliability of operation for such type of guns is very important today. For example, researching the technological processes of modern electronic industry for possibility of implementation of electron-beam technologies with using HVGDEG's, is the complex scientific and engineering problem, and its solving can stimulate and accelerate the development of this advanced industry today [13–18]. In practice, to quickly achieving of this important goal and solving this problem, the means of mathematical and computer simulation of complex and self-maintained physical processes in HVGD is generally used [8]. Mostly the self-maintained optic and energetic of HVGD in such models have been studied carefully today [8, 22, 23].

However, the operation temperature of the cathode surface, with taking into account, that it heating by the bombarding by accelerated ions and neutral particles and cooling by the cool liquid, was not thoroughly investigated. Generally, this important gun parameter was not estimated theoretically and was not measurement. But, in any case, heating of cathode surface is influence significantly to the stability of the HVGDEG operation [7, 8]. By this reason, the different constructions of cathode cooling systems for HVGD guns with different power are used in practice today. Therefore, considering the structure of used cooling systems, as well as analytical and numerical methods of its simulation, is the aim of this chapter.

For the simulation the thermodynamic processes at the cathode surface such different approaches have been used [24]:

1. Simple approach. It based on analyzing the structure and geometry of used cooling system and solving the thermodynamic problem in stationary mode by solving the Boltzmann Thermodynamic Equation (BTE).
2. A more sophisticated approach. Considering of nuclear of heat Partial Differential Equation (PDE) and finding the expansion to functional rows. Mostly integral exponential and Bessel functions are used to forming such solutions.
3. A most sophisticated approach. Numerical solving of heat PDE with using corresponded computer software.

It has to be pointed out, that analyzing of different theoretical approaches for defining the cathode temperature also can be used for simulation constructions of electron guns with heated cathode, therefore proposed in this chapter review of solving the heat PDE can be considered as universal.

The structure of chapter is follows. At the next, third part, the general approaches for solving heat PDE are considered and main equations are obtained and given. At the fourth part the structures of cooling systems in HVGDEG of different power and its' geometry parameters are presented. At the fifth part the presented constructions are simulated with using simple analytical relation, based on BTE. Experimental estimations of the cathode energetic efficiency for different cathode material and sort

of residual gas are also given in this section. At the sixth part analyzing the temperature distribution at the cathode surface with using Bessel functions and integral exponential function is considered. At the seventh part the temperature distribution at the cathode surface is analyzed numerically with using Solidworks CAM system. Describing of basic structure of Solidworks system and its' options is also given here. And finally, at the eighth section, obtained simulation results are discussed and recommendations for choosing cooling system depend on gun power are also given.

## 3 Heat Partial Differential Equation and Basic Methods of Its' Solving

In the general form heat PDE is written as follows [24]:

$$\frac{\partial T(x,y,z,t)}{\partial t} - a^2\left(\frac{\partial^2 T(x,y,z,t)}{\partial x^2} + \frac{\partial^2 T(x,y,z,t)}{\partial y^2} + \frac{\partial^2 T(x,y,z,t)}{\partial z^2}\right) = 0;\ a = \frac{\lambda}{c\rho}, \tag{1}$$

where $x$, $y$, $z$—the spatial coordinates, $t$—time, $\lambda$—thermal conductivity of considered part of construction, $a$—coefficient of thermal diffusion, $c$—isobaric heat capacity, and $\rho$—weight density of material.

The main methods of solving the PDE thermodynamic Eq. (1) are the following [2, 24–26]:

1. Using of numerical methods for solving the Eq. (1) with defined boundary and initial conditions. For example, well-known numerical Crank—Nicolson method can be used. For the simple one-dimensional case the corresponded finite-difference equation is written as follows [25, 26]:

$$\frac{T_{j,n+1} - T_{j,n}}{k} = a^2 \frac{T_{j-1,n+1} - 2T_{j,n+1} + T_{j+1,n+1} + T_{j-1,n} - 2T_{j,n} + T_{j+1,n}}{2h^2}, \tag{2}$$

where $k = \Delta t$—step of integration by the time, $h = \Delta x$—step of integration by the spatial coordinate $x$, $j$—number of current node at the spatial coordinate $x$, $n$—number of current node at the time $t$. It should be pointed out, that finite-difference scheme of Crank—Nicolson method, represented by the set of Eq. (2), is implicit, therefore the values $T_{j,n}$ are calculated by solving of three-diagonal linear system:

$$AT = B, \tag{3}$$

where $A$ and $B$—are the three-diagonal sparse matrix and the vector, defined by the set of finite-difference Eq. (2), $T$—vector of defined temperature values $T_{j,n}$.

Numerical methods, based on solving the set of linear Eqs. (2, 3), are usually given the high precision of solution, but, in any case, they are generally very sophisticated [25, 26]. Usually for numerical solving the Eq. (1) with using relations (2, 3) applied

the CAD and CAM systems, for example, such as MatLab [25] or Solidworks [27, 28]. Examples of using Solidworks system for simulation of temperature distribution at the HVGDEG cathode surface will be presented at the section 7. Describing the basic options and functions of Solidworks CAM system is also given at this section of chapter.

2. Analytical solving of Eq. (1) with considering its nuclear, which basically is written as follow function [2, 24]:

$$\Phi(x,t) = \frac{1}{\left(2a\sqrt{\pi t}\right)^n} e^{-\frac{|x|}{4a^2 t}}, \tag{4}$$

where $n$—number of current basic function.

This method has some advantages relatively to the first one, because obtained analytical solution can be analyzed for finding the optimal structure of cooling system. But, in any case, forming of functional rows with using Eq. (4) is also sophisticated task [24]. In additional, convergence of such functional rows usually isn't guaranteed [29]. For solving the thermodynamic problems with using Eq. (4) enormous MATLAB libraries of mathematical functions also can be used [25]. Corresponded example for calculation the temperature of HVGDEG cathode surface, in which on the base Eq. (4) formed the analytical solution with using integral exponential and Bessel functions, will be presented below at the section 6 of this chapter.

3. Excluding the time parameter from Eq. (1) and simplifying in to the BTE for considering the stationary thermodynamic task [2, 24]. In the general form for cooling systems, considered in the part 2, BTE can be written as follows [24]:

$$P_c = \frac{S_c(T_c - T_l)}{\sum\limits_{i=1}^{n} \frac{l_i(T)}{\lambda_i(T)}}, \tag{5}$$

where $P_c$—the power of HVGD, dissipated at the cathode surface, $S_c$—the square of cooled surface, $T_c$—the temperature of cathode surface, $T_l$—temperature of cooled liquid, $l_i$—the length considered part of construction with number $i$, $\lambda_i$—thermal conductivity of this part of construction, $n$—total number of the construction parts. Eq. (5) can be rewritten as explicit dependence $T_c(P_c, T_l)$ as follows:

$$T_c = P_c \frac{\sum\limits_{i=1}^{n} \frac{l_i(T)}{\lambda_i(T)}}{S_c} + T_l. \tag{6}$$

The analytical solution of Eq. (6) for the basic constructions of cooling system of HVGDEG, will be considered at the fifth part of this chapter. At the next part the basic constructions of cooling system and its' geometry parameters will be considered.

## 4 The Basic Constructions of the Cathode Cooling System of High Voltage Glow Discharge Electron Guns

Three standard constructions of cathode cooling items are presented at Fig. 2. There are the following:

1. The systems with massive base without forced cooling. The basic structure of such system is presented at Fig. 2a.
2. The systems with forced cooling through the base. The basic structure of such system is presented at Fig. 2b.
3. The systems with direct cooling of the cathode body. The basic structure of such system is presented at Fig. 2c.

It has to be pointed out, that on a constructive point of view the first cooling system is the simplest one, as well as the third is the most sophisticated. Therefore, using of sophisticated cooling systems is generally lead to complication of gun design, and in the such case the cost of developed electron gun is increased. The service of such gun also became complicated, and in such conditions the price of the guns with direct cooling of the cathode body is generally increased. But the main technical problem of providing the HVGD aluminum cathodes operation is maintaining the temperature of its surface in the range 150–200 °C [7, 8].

As it is clear from the Fig. 2, the main geometry parameters of considered constructions are the following: $l_c$—the cathode thickness at the symmetry axis; $l_b$—the length of base; $l_g$—the length of gap between the cathode and base; $R$—transversal radius of cooling system.

The simple analytical relation for calculation the cathode surface temperature for presented at Fig. 2 basic construction of cooling systems will be considered at the next part of chapter. All these relations are based on relation (6), presented at the part 3.

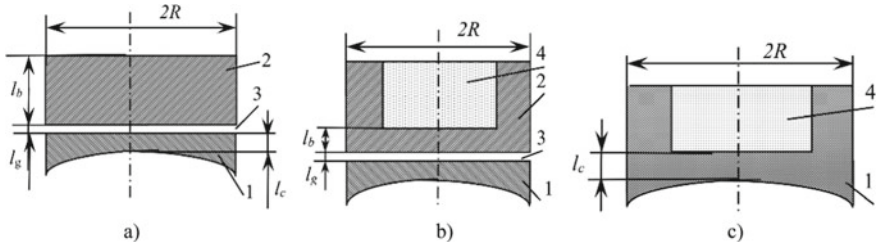



**Fig. 2** General structure of the systems of cathode cooling with the base constructions: **a** of first type, **b** second type and **c** third type. Where: 1—cathode, 2—base, 3—gap between the cathode and the base, 4—cooling liquid

## 5 Simple Analytical Relations for Estimation the Operation Temperature of Cathode Surface and the Cathode Energetic Efficiency

Solving of Eq. (6) for considered constructions of cooling system, have been presented at Fig. 2, are given and systematized at the Table 1.

At the Table 1 in the relations (7–9) such abbreviatures and symbols are used: $P_c$—the power of HVGD, dissipated at the cathode surface, $R_c$—radius of the cathode sphere, from which the emission of electrons is take place, $T_e$—the temperature of environment, $l_g$—the average value of the gap between the cathode and the base, $\lambda_b$—the thermal conductivity of the base material, $\lambda_c$—the thermal conductivity of the cathode material, $\lambda_g$—the thermal conductivity of discharge gap with taking into account the type of operation gas and its pressure in the gun volume, $\gamma$—the secondary ion-emission coefficient [7], $\alpha_l$—coefficient of heat transfer of coolant through the base, $T_l$—the temperature of coolant, $v_l$—the velocity of liquid flow, m/s, $k_1$ and $k_2$—the semiempirical coefficients. For example, for water at the normal conditions, the coefficients $k_1$ and $k_2$ are defined by the values $k_1 \approx 350$, $k_2 \approx 21{,}000$ [30, 31]. Main geometrical parameters of electrodes systems, presented at Fig. 2, are defined by the self-maintained optics of high voltage glow discharge, described in [8]. Choosing of suitable cathode material is motivated by its' physical properties, including high value of the secondary ion-emission coefficient $\gamma$ and the thermal conductivity of the cathode material $\lambda_c$ [30, 31]. Most often, this choice is of a contradictory and compromise nature, however, the practice of using cathode materials has shown that aluminum cathodes have the best emission properties with good thermal conductivity. It is proved theoretically in the theory of HVGD, that energetic efficiency of the cathode material $H_c$ is defined by the following relation [32–34]:

**Table 1** Analytical relations for calculation the average cathode surface temperature, have been obtained by solving Eq. (6) for the different type of cooling systems, which basic structures are presented at Fig. 2

| Type of construction | Analytical relations |
|---|---|
| 1. The systems with massive base without forced cooling, Fig. 2a | $T_c = \frac{\gamma P_c \left( \left( \frac{R_c + l_c - \sqrt{R_c^2 - R^2}}{\lambda_c} \right) + \frac{l_g}{\lambda_g} + \frac{l_b}{\lambda_b} \right)}{R_c^2 (1+\gamma) \left( \arcsin\left( \frac{R}{2R_c} \right) - \frac{R}{2R_c} \right)} + T_e$, (7) |
| 2. The systems with forced cooling through the base, Fig. 2b | $T_c = \frac{\gamma P_c \left( \left( \frac{R_c + l_c - \sqrt{R_c^2 - R^2}}{\lambda_c} \right) + \frac{l_g}{\lambda_g} + \frac{l_b}{\lambda_b} \right)}{\alpha_l R_l^2 (1+\gamma) \left( \arcsin\left( \frac{R}{2R_c} \right) - \frac{R}{2R_c} \right)} + T_l,$ $\alpha_l = k_1 + k_2 \sqrt{v_l},$ (8) |
| 3. The systems with direct cooling of the cathode body, Fig. 2c | $T_c = \frac{\gamma P_c \left( R_c + l_c - \sqrt{R_c^2 - R^2} \right)}{\lambda_c R_c^2 (1+\gamma) \alpha_l \left( \arcsin\left( \frac{R}{2R_c} \right) - \frac{R}{2R_c} \right)} + T_l$. (9) |

$$H_c = \gamma\lambda_c = \frac{I_{ec}}{W_c} = \frac{d(\gamma U_{ac})}{dU_{ac}}, \tag{10}$$

where, $I_{ec}$—emission current of the cathode, $W_c$—the power, dissipated at the cathode surface, $U_{ac}$—acceleration voltage.

For the aluminum as cathode material parameter $H_c$ have the maximal value, and this effect caused by the presents of thin aluminum oxide film at the surface of cathode [32–34]. In the real HVGD systems the value $H_c$ is depended on cathode material, sort of residual gas, as well as on acceleration voltage [32–34]. The experimental dependences of the cathode energetic efficiency $H_c$ on the HVGD current $I_d$ for different metals, gases and acceleration voltage are presented at Fig. 3.

The results of estimation the cathode surface temperature, have been obtained with using analytical relations (7–9), are presented at Fig. 4. The calculations were provided at the MATLAB software [25] for such general parameters of cooling system: $\gamma = 3.4$; $R_c = 0.8$ m; $R = 0.35$ m; $l_c = 0.01$ m; $l_g = 0.003$ m and $l_a = 10^{-5}$ m, aluminum as cathode material, cooper as base material, water as coolant and nitrogen with mixing 1% of oxygen as operation gas.

As it is clear from dependences, presented at Fig. 3, that the best energetic efficiency given the cathodes from $LaB_6$. But both at the theoretical and experimental

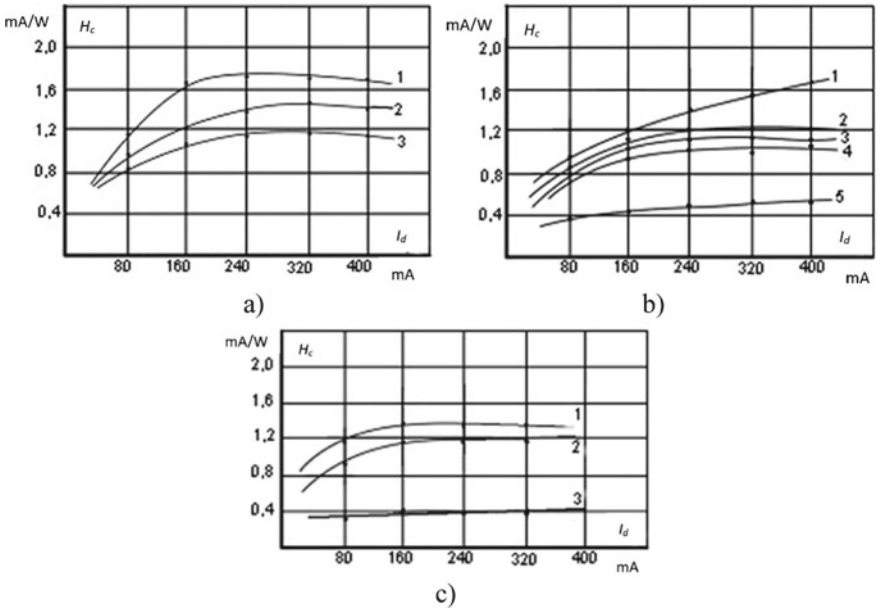



**Fig. 3** Dependence of the energy efficiency of cold cathodes on the discharge parameters: **a** for various accelerating voltages: 1–15 kV, 2–20 kV, 3–25 kV, operation gas—air; **b** for various cathode materials: 1—$LaB_6$, 2—Al, 3—Al + 10% Cu, 4—Al + 20% Cu, 5—stainless steel, accelerating voltage—20 kV; **c** for various operation gases: 1—helium, 2—air, 3—argon, accelerating voltage—25 kV, cathode material—aluminum

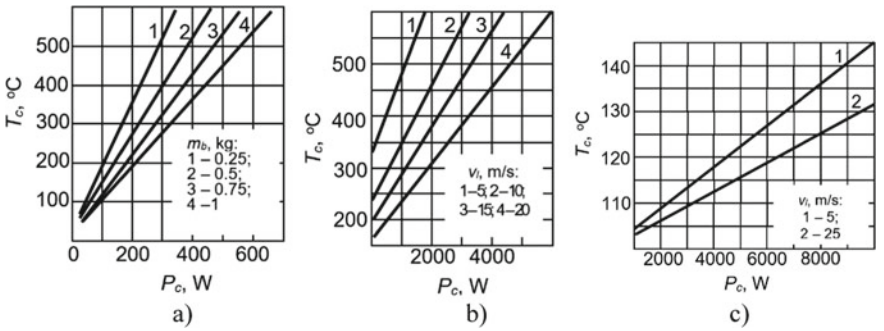



**Fig. 4** Dependences of the operation cathode temperature on the based construction of cooling system and its thermodynamic parameters: **a** the first type of construction (Fig. 2a); **b** the second type (Fig. 2b); **c** the third one (Fig. 2c)

point of view the main disadvantage of such cathode the small value of thermal conductivity, therefore effective cooling of this type of cathode is impossible. After the $LaB_6$, as it is clear from Fig. 3b, the maximal value of energetic efficiency giving the cathodes from pure aluminum. With including the cooper at the aluminum alloy, the energetic efficiency is reduced, and it can be explained by reducing the secondary ion-emission coefficient $\gamma$ [32–34]. But the advantage of AlCu alloys as the cathode material in HVGDEGs is the higher thermal conductivity correspondently to pure aluminum, therefore the problem of effective cooling of such cathodes is generally simplified. The worse value of energetic efficiency is observed for the steel, it caused by the low level both the secondary ion-emission coefficient and thermal conductivity for this material [31, 32].

The high level of the energy efficiency of the cathodes in the HVGDEGs is also influenced by the choice of the suitable operation gas. It was theoretically and experimentally substantiated that the greatest efficiency of the cathode operation is provided by light gases, like hydrogen or helium, in which the coefficient of electrical transfer ($K \approx \gamma/S_c$) is the maximal value. Since the aluminum cathode has a high coefficient of ion-electron emission due to the oxide film on its' surface, which is destroyed under the influence of ion bombardment, especially in the regime of high discharge powers, it is necessary, for films' restoring, to add at the mix of operation gases the small amount of oxygen, up to 1%. At an emission current density not exceeding 0.1 A/cm$^2$, a sufficiently stable discharge combustion is ensured at a low sputtering rate of the cathode, which will make it possible to stabilize its emission characteristics and increase the service life [32–34].

Corresponded experimental dependences of the cathode energetic efficiency on the sort of operation gas are presented at Fig. 3c. Reducing of the cathode energetic efficiency with increasing of acceleration voltage $U_{ac}$, observed at the Fig. 3a, is caused by increasing the electrical transfer coefficient $K$ because of increasing the effectivity of sputtering the oxide film at the aluminum surface.

Analyze of the obtained simulation results, given at the Fig. 4, will be provided at the eighth section of the chapter.

## 6 Simulation of Temperature Distribution at the Cathode Surface with Using the Set of Specific Functions

Solving of heat PDE (1) with considering its nuclear, which is generally based on using function (4) and expansion to functional series, is very sophisticated problem and such mathematical task can be solved only to very simple boundary conditions [2, 24]. For example, for the schemes of cooling systems, presented at the Fig. 2, this task can be solved only for most simple system with cooling through massive base, presented at Fig. 2a, but with strong restrictions to the geometry of this system, which will be considered at this section below.

But, in another side, this cooling system can be significantly improved, taking into account the particularities of HVGD and physical conditions, in which operated the cathode [7, 32–34]. In the conditions of HVGD combustion, almost the entire surface of the cathode is bombarded with positive ions, however, the ion current density at the cathode is unevenly distributed and over the peripheral part of the cathode it is significantly less. In such conditions, at a low temperature of the peripheral region of the cathode, a positive charge of ions is always accumulated in the dielectric $Al_2O_3$ film. Namely this effect is lead to a change in the conditions of ionization and to the initiation in the near-cathode region high-current microarcs. As a result, the discharge current is sharply increased and stability of HVGD lighting is significantly decreased [7, 32–34]. On the other hand, when the temperature of the central region of the cathode rises, the coefficient of secondary ion–electron emission and the cathode efficiency coefficient $H_c$ decreases. In this regard, in HVGDEG, to ensure the normal operation of the cold cathode and increase the stability of the discharge combustion, it is advisable to cool effectively through the base namely the central region of the cathode and to maintain a higher temperature on its peripheral region. The corresponded construction of the cooling system, which makes it possible to provide uneven cooling of the central region of the cathode surface through the base, is shown in Fig. 5. This construction generally can be considered as modification and improving of construction, presented above at Fig. 2a.

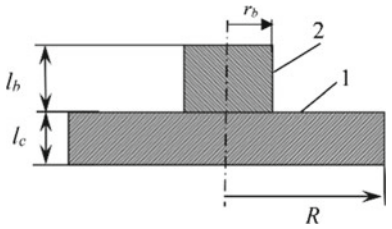



**Fig. 5** Cooling system of the cathode through the base for providing the uneven temperature distribution on its surface: 1—cathode, 2—massive copper base, $r_b$—radius of the base

Considering the formulated simulation problem, it is necessary to take into account that the constrictive scheme, presented in Fig. 5, in comparison with the constrictive scheme, presented in Fig. 2a, is significantly simplified. Namely, the gap between the cathode body and the base isn't take into account, because analytical solution of Eq. (1) for the system with gap isn't existed. Such approach is plausible only for the close contact between the cathode and cooper body surfaces with the length of gap range of $l_g < 10^{-6}$ m. The brazing of cathode to body is also possible. Another limitation, that the analytical solution can't be obtained for the spheric cathode surface, only plane cathode geometry can be considered.

For the boundary conditions determined by the geometry of the cooling system presented in Fig. 5, when the constraints $l_c << r$ and $l_c << R$ are satisfied, the heat PDE (1) with taking into account the relation (4) has an analytical solution, which is written as the following set of functional relations [24]:

$$
\begin{gathered}
T_{\max} = \frac{W_c\left(\mathrm{Ei}\left(-kr^2\right)-\mathrm{Ei}\left(-kR^2\right)-2\exp\left(-kR^2\right)\ln(R)\right)}{4\pi k\lambda_c R^2} + \\
+\frac{W_c l_c\left(\exp\left(-kr^2\right)-1\right)}{\mathrm{ap}R^2} + \frac{W_c\exp\left(-kR^2\right)\ln(R)}{2k\lambda_c\pi R^2} + A\left(\mathrm{I}_0\left(r\sqrt{\frac{\alpha}{\lambda_x l_x}}\right)-1\right) + T_{\min}, \\
T_{\min} = A + \frac{W_c l_c}{\pi R^2} + T_0, \\
A = \frac{\sqrt{\frac{\lambda_c l_c}{\alpha}}\left(\frac{W_c\left(\exp\left(-kr^2\right)-\exp\left(-kR^2\right)\right)}{2k\lambda_c\pi R^2 r}+\frac{2W_c krl\exp\left(-kr^2\right)}{\mathrm{ap}R^2}\right)}{\mathrm{I}_1\left(r\sqrt{\frac{\alpha}{\lambda_c l_c}}\right)}, \\
\mathrm{Ei}(-x) = \int_{-\infty}^{-x}\frac{\exp(t)}{t}dt,\ \mathrm{I}_v(z) = \left(\frac{z}{2}\right)^v\sum_{k=0}^{\infty}\frac{\left(\frac{z^2}{4}\right)^k}{k!\Gamma(v+k+1)},\ \Gamma(a) = \int_0^{\infty}\mathrm{e}^{-t}t^{a-1}dt,
\end{gathered}
\tag{11}
$$

where $T_{\max}$—the maximum temperature at the periphery of the cathode, $T_{\min}$—the minimum temperature in the central region of the cathode, $\lambda_c$—the thermal conductivity of the cathode material, $\mathrm{Ei}(-x)$ —integral exponential function, $\mathrm{I}_0(x)$ and $\mathrm{I}_1(x)$ —modified Bessel functions of the first kind of zero and first order, $\Gamma(a)$ — Euler's gamma-function [29], $k = 0.2$ cm$^{-2}$—the heat flow concentration coefficient.

Solving of the set of Eq. (11) have been realized at the MATLAB software, and obtained graphic dependences are presented at Fig. 6. The calculations were carried out for the following values of geometric parameters of cathode cooling system, presented at Fig. 5: $R = 0.35$ m, $l_c = 0.01$ m, $l_b = 0.1$ m, the cathode material is aluminum.

It is clear from obtained simulation results, that, depend on the radius of base, the difference between the temperature at the center and at the peripheral region of the cathode can be range of 250 °C and more, depend on dissipated power. Further investigation of stability of operation the HVGDEGs with cooling the central region of cathode surface though the massive base is the separate engineering problem.

Today in the powerful electron guns, which operated in uninterrupted regime, always applied the systems with forced cooling, like presented at Fig. 2c [6, 33]. Furthermore, using of electrodes systems with plane cathode surface isn't preferable for forming the intensive elector beams with high power density, because using the sphere geometry of cathode surface provides the focusing of beam electrons at the near-cathode region. In any case, preliminary theoretical estimations with using

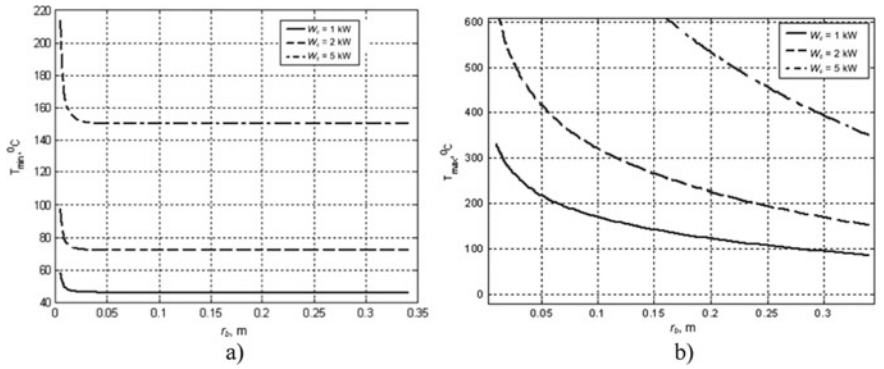



**Fig. 6** Dependence of a) minimum and b) maximum temperature of cathode surface on the radius of base $r_b$ for cooling system, presented at Fig. 5, obtained with using set of relations (11)

relations (11) are very interesting for study the additional possibilities of further improving the constructions of high power HVGDEG's in order to increase the stability of its' operation.

## 7 Using Solidworks CAM System for Simulation the Temperature Operation Regime of Cold Cathode of Powerful High Voltage Glow Discharge Electron Guns

Solidworks CAM system is well-known and widespread package that allows to create spatial models of construction parts based on drawings of corresponded primitives. Simulation of mechanic stresses and the temperature regimes on different details in Solidworks software is also possible. Also, this system has a powerful algorithm for calculating loads of different nature. Based on the physical parameters of the materials, which are available in the program data bases and include almost all widely used metals, alloys, ceramics, plastics, liquids, gases and other materials, mechanical loads and deformations, as well as thermal loads can be simulated. The aerodynamics or hydrodynamics of an object in a liquid or gas flow also can be analyzed [27, 28]. Therefore, Solidworks CAM system can be generally considered as universal multipurpose simulation software instrument [27, 28].

For calculation the temperature distribution the numerical method, based on Crank—Nicolson algorithm and set of Eqs. (2, 3) is used in Solidworks package [27, 28].

To get started, open the program and create a new project is necessary. For example, for creating the model of separate simple detail the menu function "Gost Part" is available [27, 28].

The next step is to create a draft of the required simple part of construction, for example, a fitting with an external thread. After previously analyzed the shape of

this part, one can conclude, that it is a surface of rotation, with an external thread and a hexagonal slot. Taking this into account, one can to build a small figure with the axial section, similar to fitting. All required dimensions are marked at the draft, so, the next step is to rotate it around the axis by clicking on the command "Rotated Base". After that, in the dialog box, selected the axis of rotation, the closed area, and the rotation angle [27, 28]. After these actions, the program Solidworks will demonstrate a preliminary view of the body, which have been formed. Corresponded draws of fitting element during settings the thread parameters and the final spatial view of simulated part of construction are presented at Fig. 7.

Thus, it is possible to create easily in the Solidworks program spatial models of rather difficult details, and then to analyze features of its operation in real designs [27, 28]. To continue the simulation process material of part, which have been drafting, also have to be specified at the dialog box. Best material for the fitting element is brass.

At the similar way in Solidworks formed the draft of other details. The spatial view of cathode and its' holder are presented at Fig. 8. Cathode radius $R$ is 44.5 mm and radius of cathode sphere $R_{sp}$ is 100 mm. These geometry parameters of gun construction are defined by the power of electron beam, which contain 100 kW, and by the position of focus of electron beam [7, 33].

The draft of assembled cathode item, obtained with using Solidworks CAM software, is presented at Fig. 9.

To simulation the cathode operating temperature an application of the Solidwork software Flow Simulation package was used [27, 28]. This program allows to simulate the flow of liquids or gases within certain geometric limits, taking into account the heat exchange between bodies, liquids and interacting gases. Since this program is primarily focused on the development of spatial models, it performs calculations




**Fig. 7** The testing draft of fitting element, obtained at the Solidworks CAM software: **a** during setting the thread parameters; **b** final spatial view




**Fig. 8** The final spatial view of the cathode **a** and its' holder **b**, obtained with using Solidworks CAM software

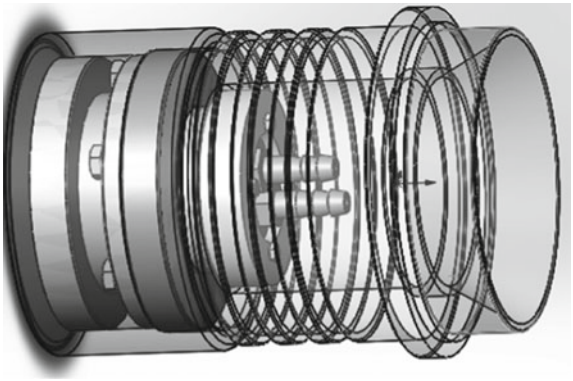


**Fig. 9** The final spatial view of assembled cathode item at the Solidworks CAM software

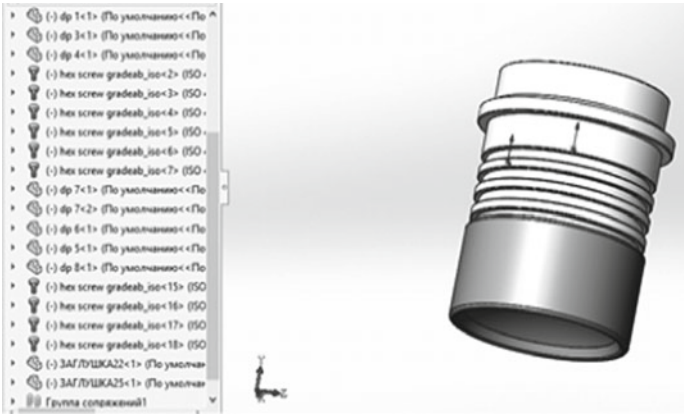


**Fig. 10** Choosing option "Flow Simulation" at the Solidworks CAM software for assembled cathode item

of physical parameters with a small certain error, which does not allow to obtain a completely accurate result. However, the accuracy of the calculations is usually enough to make the general conclusions about the correctness of applying the certain cooling system under the preliminary study [27, 28].

To simulation the temperature distribution, one need to have an assembly or part in the form of a spatial model, which must be created in the previous stage. This stage has been generally described above. Then one need to open the project file and choose the option "Flow Simulation", like shown at Fig. 10 [27, 28]. At the next step is necessary to create a new project with the specified physical quantities, materials properties and initial conditions for the implementation of simulation project.

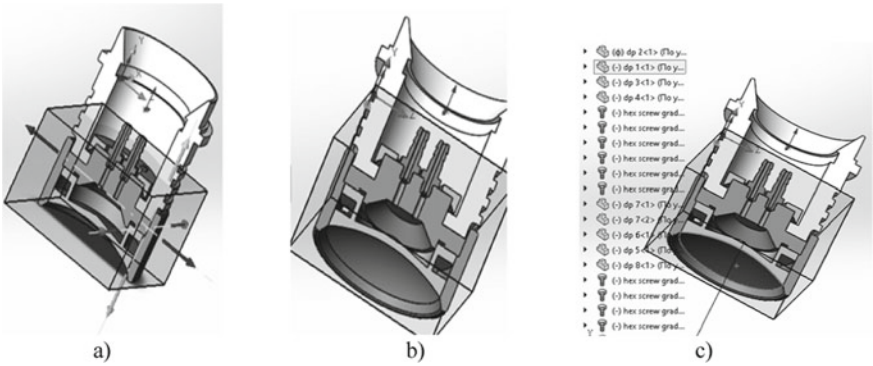



**Fig. 11** The results of performance the operations of visualization calculation area **a**, defining boundary conditions **b** and defining the source of heating **c** for simulation the cathode cooling system at "Flow Simulation" software

The next steps of simulation procedure are the follows [27, 28].

1. Visualization of calculation area, option "Calculation Area".
2. Defining of boundary conditions, option "Boundary Conditions".
3. Defining the source of heating, option "Surface source of heating".

To solve the problem of determining the temperature distribution in the cooling system of the cathode assembly, the boundary conditions have to be defined for the water, which cooling the cathode. For considered task the volumetric flow rate was choose 1 l/s, the pressure in the cooling system $5 \cdot 10^5$ Pa and the flow rate of water 20 m/s. The power, dissipated at the cathode surface, for 100 kW beam power was choose 5 kW. Such value is generally corresponded to physics of HVGD [7, 33].

Examples of providing the corresponded steps and results of its' performance in "Flow Simulation" software are shown at the Fig. 11.

The final result of simulation the temperature distribution at the cathode surface in the Solidworks CAM software is presented at Fig. 12.

Analyze of simulation results, have been obtained with using different considered approaches for solving the important thermodynamic task of effective cooling the cathode surface at powerful HVGDEG's, will be provided at the next section of this chapter.

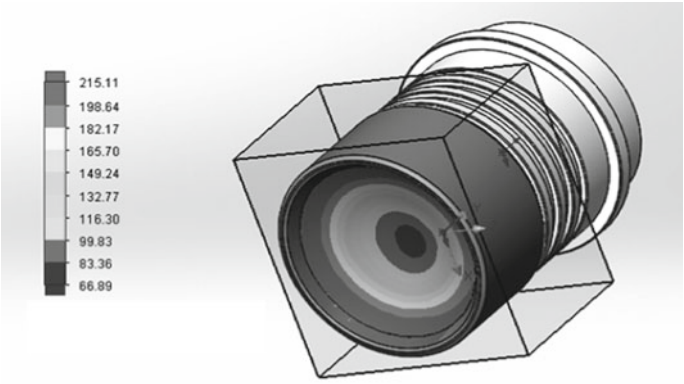



**Fig. 12** The result of simulation the temperature distribution at the cathode surface of 100 kW HVGDEG, obtained with using Solidworks CAM software

## 8 Analyze of Obtained Simulation Results, Recommendations and Discussion

Obtained simulation results for preliminary and precisions estimation of the cathode surface temperature, presented at Figs. 4, 6 and 12, are very important both theoretically and practically for designing the new generation of powerful HVGDEG. Generally, these results allow to formulate useful practice recommendation for experts in designing of electron-beam technological equipment relatively to possibility of using in HVGD electron sources the different basic constructions of cooling systems, presented at Fig. 2. The choosing of suitable construction of cathode cooling block is usually the controversial-compromise engineering and economy task between the correspondence to requirement of providing the stability of operation of electron gun as a part of technological equipment, cheapness of gun and simplicity of its' exploitation.

Therefore, the type of applying cooling system is strongly depend on guns' power and its technological possibilities. The main recommendations for choosing the type of cooling system are the following:

1. For the guns with relatively low power, range of 5–10 kW, and especially for the guns, predestined to realizing the short-time technological operations in pulse regime, like single-impulse welding of small-sized and thin-wall items [6, 32], using of the system without forced cooling, which basic scheme of construction is presented at Fig. 2a, is generally preferable. Such solution allows significantly simplify the construction of electron gun, as well as its exploitation, therefore the price of such HVGD electron gun and corresponding exploitation costs are generally getting lower. The simulation results are shown, that overhead of cathode surface in such energetic regime of guns' operation isn't observed.

The same conclusion is also given by the many years of practice of designing and exportation of technological HVGDEGs.

2. For the guns with the average power, range of 10–30 kW, using of the system with forced cooling through the base, which basic structure is presented at Fig. 2b, is generally preferable. In constructions of such cooling systems providing of reliable contact between the cathode body and base by reducing of vacuum gap distance $l_g$ is generally necessary. Corresponding to simulation results, recommended value of this geometrical parameter of cathode cooling system construction is $l_g < 10^{-6}$ m, therefore mechanical treatment of the adjoined surfaces of cooling body and inner cathode plane should be strongly accurate. Such cooling systems also provided the simplicity of electron gun exploitation and repairing, because necessary changing of the spared cathode can be provided without depressurization of liquid volume of cooled base. The HVGDEGs with such type of cooling system construction can be successfully applied in the technologies of deposition thin dielectric films in electronics devices production [9, 12–18]. Using such type of guns in the metallurgy for obtaining pure metals and ceramics ingot is also possible [3].
3. For the HVGD electron guns with the great power, range of 40–500 kW, applying of systems with direct cooling of the cathode body, presented at Fig. 2c. is generally necessary, because the construction of such type has to provide the maximal extraction of thermal power from the cathode body. But the cost of gun and the expenses to its' exploitation are significantly grows. High voltage glow discharge electron guns of such type can be successfully applied in metallurgy for high-pure refining refractory metals and ceramic materials [3].
4. For designed HVGDEG's the energetic efficiency of cathode operation can be roughly estimated with using Eq. (10). Usually, the value of cathode energetic efficiency $H_c$ is in the range 0.4–1.5 mA/W.
5. The simulation results, obtained with using simple analytical relations (7–9), are generally similar to results, obtained with using Solidworks CAM system. Difference between the results of analytical and numerical simulation were in the range 10–15%. Corresponded simulation results are presented at Fig. 4 and Fig. 12. Therefore, using of simple analytical relations (7–9) for the preliminary estimation of average cathode surface temperature on the first step of gun designing is generally recommended. Obtained analytical relation (8, 9) are also allows to estimate the temperature of coolant and its velocity for obtaining the temperature of aluminum cathode in the range of 150–200 °C.
6. Researching of uneven cooling of the central region of the cathode surface through the base, like at the construction, presented at Fig. 5, with using special functions and analytical Eqs. (11), can give the further impulse to improving the constructions of powerful HVGDEG. But the limitations to the geometry of cooling system, which can be simulated analytically with using special functions, are really very strong, and the plane cathode geometry generally not correspond to the HVGD self-maintained electron-ion optics by the required focal beam parameters and beam power density [7]. Dependences of maximal and minimal

temperature of cathode on the dissipated power and radius of cooling base are presented on Fig. 6.

7. Using of Solidworks CAM computer software allow to estimate preliminary the temperature distribution at the cathode surface for certain considered construction of cooling system. The main advantage of this simulation software is possibility of forming complex construction from the simple mechanical elements. The thermodynamic properties of used materials and necessary coefficients also can be exported from original Solidworks data bases. In any case, the precision of rough estimations, have been made with applying Solidworks CAM software, is generally enough at the preliminary step of HVGDEGs designing.
8. For adjusted and optimized constructions of cooling systems for HVGDEGs with beam power up to 500 kW the temperature of aluminum cathode surface is maintained in the range 150–250 °C.

## 9 Conclusion

Presented in the chapter analytical relations (7–9) and simulation results, given at the Figs. 3, 6 and 12 correspondently, allows to estimate preliminary the cathode surface temperature at the HVGDEGs, formed electron beams with different power. For providing the effective cooling of the cathode body and obtaining at the cathode surface optimal operation temperature range of 150–250 °C, different basic constructions of cooling system have been proposed. Recommendations for choosing the type of cooling system, depending on the gun's power, are formulated at the section 8. The energetic efficiency of cathodes from different materials have been investigated experimentally. The methods of calculation the average cathode surface temperature as well as obtained results, presented in chapter, may be interesting for the experts in designing and exploitation of modern electron-beam industrial equipment and corresponded technologies. Possible application of HVGDEGs in production of electronic components are also pointed out.

## References

1. Shiller S, Geisig U, Panzer S (1980) Elektronno-luchevye technologiyi [Electron-beam technology]. Energiya, Moscow (in Russian)
2. Ryikalin NN, Zuev IV, Uglov AA (1978) Osnovyi elektronno-luchevoy obrabotki materialov [Foundation of Electron-Beam Material Treatment]. Mashinostroenie, Moscow (in Russian)
3. Ladokhin SV, Levitskiy NI, Chernyavskiy VB, et al (2007) Electronno-luchevaya plavka v liteinom proizvodstve [Electron beam melting in the foundry production]. Stal, Kyiv. (in Russian)
4. Sliva A, Dragunov V, Goncharov A, Shcherbakov A (2012) Electron beam welding method in the magnetic field. Electrotech Electron (E+E) **47**(5–6):128–133
5. Krasik YE, Gleizer JZ, Krokhmal A, Chirko K, et al (2003) High-current electron sources based on gaseous discharges. Vacuum 77(4):391–398

6. Gruzdev VA, Zalesski VG, Antonovich DA, Golubev VP (2003) Universal plasma electron source. Vacuum 77(4):399–406
7. Denbnovetskiy S, Melnyk V, Tugai B, Tuhai S, Wojcik W, Lawicki T, Assambay A, Luganskaya S (2017) Principles of operation of high voltage glow discharge electron guns and particularities of its technological application. In: Proceedings of SPIE. The International Society of Optical Engineering, pp 10445–10455
8. Melnyk I, Tuhai S, Pochynok A (2021) Universal complex model for estimation the beam current density of high voltage glow discharge electron guns. In: Ilchenko M, Uryvsky L, Globa L (eds) Lecture notes in networks and systems, vol 152, pp 319–341. https://www.springer.com/gp/book/9783030583583
9. Grechanyuk MI, Melnyk AG, Grechanyuk IM, et al (2014) Modern electron beam technologies and equipment for melting and physical vapor deposition of different materials. Electrotech Electron (E+E) 49(5–6):115–121
10. Hassel T, Beniyash N, Murray N, Konya R, Bach Fr-W (2012) Non-vacuum electron beam for cutting application. Electrotech Electron (E+E) 47(5–6):146–151
11. Vassilieva V, Vutova K, Donchev V (2012) Recycling of alloy steel by electron beam melting. Electrotech Electron (E+E) 47(5–6):142–145
12. Mattausch G, Zimmermann B, Fietzke F, Heinss JP, Graffel B, Winkler F, Roegner FH, Metzner C (2014) Gas discharge electron sources – proven and novel tools for thin-film technologies. Electrotech Electron (E+E) 49(5–6):183–195
13. Druzhinin AA, Ostrovskii IP, Khoverko YN, Liakh-Kaguy NS, Vuytsyk AM (2014) Low temperature characteristics of germanium whiskers. Funct Mater 21(2):130–136
14. Druzhinin AA, Bolshakova IA, Ostrovskii IP, Khoverko YN, Liakh-Kaguy NS (2015) Low temperature magnetoresistance of InSb whiskers. Mater Sci Semiconductor Process 40:550–555
15. Zakharov A, Rozenko S, Litvintsev S, Ilchenko M (2020) Trisection bandpass filter with mixed cross-coupling and different paths for signal propagation. IEEE Microw Wirel Compon Lett 30(1):12–15
16. Zakharov A, Litvintsev S, Ilchenko M (2019) Trisection bandpass filters with all mixed couplings. IEEE Microw Wirel Compon Lett 29(9):592–594
17. Zakharov A, Rozenko S, Ilchenko M (2019) Varactor-tuned microstrip bandpass filter with loop hairpin and combine resonators. IEEE Trans Circ Syst II Exp Briefs 66(6):953–957
18. Zakharov A, Litvintsev S, Ilchenko M (2020) Transmission line tunable resonators with intersecting resonance regions. Trans Circ Syst II Exp Briefs 67(4):660–664
19. Zhang J, Yang Y, Cao S, Kovalchuk D, Wu F, Huang A (2020) Fine equiaxed β grains and superior tensile property in Ti–6Al–4V alloy deposited by coaxial electron beam wire feeding additive manufacturing. Acta Metallurgica Sinica (English Letters) 33(10):1311–1320
20. Melnyk IV, Pochynok AV (2019) Modeling of electron sources for high voltage glow discharge forming profiled electron beams. Radioelectron Commun Syst 62(6(684)):311–323
21. Denbnovetsky SV, Melnyk VI, Melnyk IV, Tugay BA (2003) Model of control of glow discharge electron gun current for microelectronics production applications. In: Proceedings of SPIE sixth international conference on material science and material properties for infrared optoelectronics, vol 5065, pp 64–76
22. Melnyk IV, Melnyk VG, Tugai BA, Tuhai SB, Mieshkova NI, Pochynok AV (2019) Simplified universal analytical model for defining of plasma boundary position in the glow discharge electron guns for forming conic hollow electron beam. In: IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO), Ukraine. Conference proceedings, Kyiv, pp 548–552
23. Melnyk IV, Tuhai SB (2012) Analytical calculations of anode plasma position in high-voltage discharge range in case of auxiliary discharge firing. Radioelectron Commun Syst 55(11):50–59
24. Isachenko VP, Osipova VA, Sukhomel AS (1975) Teploprovodnost [Thermocondactivity]. Energoatomizdat, Moscow, (in Russian). https://www.c-o-k.ru/images/library/watermarked/cok/334/33429/961f467d6c5fc3c6a02d5a5fbff490af.pdf
25. Mathews JH, Fink KD (1998) Numerical Methods. Using Matlab. Amazon

26. Samarskiy AA, Gulin AV (1989) Chislennye metody [Numerical methods]. Nauka, Moscow (in Russian)
27. Chang KH (2021) Machining simulation using solidworks CAM. SDC
28. Schilling PJ, Shih RH (2021) Parametric modeling with solidworks. SDC
29. Abramowitz M, Stegun I (1979) Spravochnik po specialnym funktsiyam [Handbook on special functions]. Nauka, Moscow. (in Russian). https://www.twirpx.com/file/54115/
30. Kuchling H (2014) Taschenbuch der Physik. 21 Edition. Hanser Verlag. (in German)
31. Espe W (1966) Materials of high vacuum technology. Pergamon Press
32. Novikov AA (1983) Istochniki elektronov vysokovol'tnogo tleiuschego razrada [High voltage glow discharge electron sources]. Energoatomizdat, Moscow (in Russian)
33. Zavialov MA, Kreindel YuE, Novikov AA, Shanturin LP (1989) Plasmennye processesy v technologicheskih electronnyh pushkah [Plasma processes in the technological electron guns]. Atomizdat, Moscow (in Russian)
34. Denbnovetskiy S, Melnik I, Melnik V, Felba J (1997) Model of beam formation in a glow discharge electron gun with a cold cathode. Appl Surf Sci 111:288–294

# Author Index


**A**
Avdieienko, Hlib, 535, 553

**C**
Chepusenko, Yevhenii, 217
Chmielewski, Ingo, 203

**D**
Dovgyi, Stanislav, 83, 181
Drozd, Sofia, 143
Dugaev, Dmitrii, 203

**G**
Globa, Larysa, 61, 83, 119
Grebinichenko, Mariya, 119

**I**
Ilchenko, Michael, 439, 469
Ilchenko, Mykhailo, 535
Iushchenko, Maksim, 203

**J**
Junfeng, Yu, 61

**K**
Kachan, Dmitry, 277
Kaidenko, Mykola, 399
Kamaraly, Roman, 535
Karpov, Kirill, 203
Kluge, Tim, 35
Klymash, Mykhailo, 165
Kopiika, Oleh, 83, 181
Kovalchuk, Dmytro, 575
Kozlov, Oleksii, 83
Kravchuk, Irina, 377
Kravchuk, Serhii, 377, 399
Kussul, Nataliia, 143

**L**
Lemeshko, Oleksandr, 235
Lemeshko, Valentyn, 235
Lioubimova, Alexandra, 217
Litvintsev, Sergii, 499
Luntovskyy, Andriy, 165
Luzianin, Ivan, 203
Lysenko, Oleksandr, 357
Lyubymova, Nina, 217

**M**
Mankivskyi, Volodymyr, 333
Mareev, Nikolai, 277
Martynova, Kateryna, 297
Matusek, Daniel, 35
Matveev, Ivan, 203
Melnyk, Igor, 575
Melnyk, Vitaliy, 575
Mersni, Amal, 235

**N**
Naumenko, Artem, 217
Nesterenko, Mykola, 333


© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

M. Ilchenko et al. (eds.), *Progress in Advanced Information and Communication Technology and Systems*, Lecture Notes in Networks and Systems 548,
https://doi.org/10.1007/978-3-031-16368-5

Nosok, Svitlana, 101
Novikov, Valery, 357
Novogrudska, Rina, 61, 119

**P**
Persikov, Mykhailo, 235
Pinchuk, Liudmyla, 499
Piven, Oleh, 101
Poliarus, Oleksandr, 217

**R**
Romaniuk, Anton, 357
Romaniuk, Valery, 357
Romanov, Oleksander, 333
Rozenko, Sergii, 499

**S**
Schill, Alexander, 35
Shelestov, Andrii, 101, 143
Shmelkin, Ilja, 35
Shumilo, Leonid, 143
Shved, Iryna, 575
Siemens, Eduard, 203, 277
Solovskaya, Irina, 317
Springer, Thomas, 35
Strelkovska, Juliya, 317
Strelkovskaya, Irina, 257, 317
Surzhykov, Mykola, 575
Sushyn, Ihor, 357
Sylantyev, Sergiy, 143

**T**
Tsukanov, Oleg, 419
Tuhai, Serhii, 575

**U**
Uryvsky, Leonid, 297

**Y**
Yailymov, Bohdan, 101
Yailymova, Hanna, 101
Yakornov, Yevgenii, 419
Yakornov, Yevhenii, 553
Yavisya, Valery, 357
Yeremenko, Oleksandra, 235
Yevdokymenko, Maryna, 235

**Z**
Zadoienko, Bohdan, 61
Zakharov, Alexander, 439, 469
Zhivkov, Alexander, 535
Zhuk, Oleksandr, 333
Zobjack, Tim, 165
Zolotukhin, Roman, 257