

ISSN 1684-8853 (print); ISSN 2541-8610 (online)

# ИНФОРМАЦИОННО- УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

4(137)/2025

4(137)/2025

PEER REVIEWED JOURNAL

# INFORMATSIONNO- UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

**Founder**

A. Vostrikov

**Publisher**Saint Petersburg State University  
of Aerospace Instrumentation**Editor-in-Chief**

E. Krouk

Dr. Sc., Professor, Moscow, Russia

**Executive secretary**

O. Muravtsova

**Editorial Board**

S. Andreev

Dr. Sc., Tampere, Finland

V. Anisimov

Dr. Sc., Professor, Saint Petersburg, Russia

B. Bezruchko

Dr. Sc., Professor, Saratov, Russia

N. Blaunstein

Dr. Sc., Professor, Beer-Sheva, Israel

M. Buzdalov,

PhD, Researcher, Saint Petersburg, Russia

C. Christodoulou

PhD, Professor, Albuquerque, New Mexico, USA

A. Dudin

Dr. Sc., Professor, Minsk, Belarus

I. Dumer

PhD, Professor, Riverside, USA

M. Favorskaya

Dr. Sc., Professor, Krasnoyarsk, Russia

L. Fortuna

PhD, Professor, Catania, Italy

A. Fradkov

Dr. Sc., Professor, Saint Petersburg, Russia

A. Hramov

Dr. Sc., Professor, Innopolis, Russia

L. Jain

PhD, Professor, Canberra, Australia

A. Myllari

PhD, Professor, Grenada, West Indies

K. Samouylov

Dr. Sc., Professor, Moscow, Russia

J. Seberry

PhD, Professor, Wollongong, Australia

M. Sergeev

Dr. Sc., Professor, Saint Petersburg, Russia

A. Shalyto

Dr. Sc., Professor, Saint Petersburg, Russia

A. Shepeta

Dr. Sc., Professor, Saint Petersburg, Russia

Yu. Shokin

RAS Academician, Dr. Sc., Novosibirsk, Russia

A. Smirnov

Dr. Sc., Professor, Saint Petersburg, Russia

T. Sutikno

PhD, Associate Professor, Yogyakarta, Indonesia

A. Tyugashev,

Dr. Sc., Professor, Samara, Russia

Z. Yuldashev

Dr. Sc., Professor, Saint Petersburg, Russia

A. Zeifman

Dr. Sc., Professor, Vologda, Russia

**Editor:** A. Larionova**Proofreader:** T. Zvertanovskaia**Design:** M. Chernenko, Yu. Umnitsyna**Layout and composition:** Yu. Umnitsyna**Contact information**The Editorial and Publishing Center, SUAI  
67A, Bol'shaya Morskaya, 190000, Saint Petersburg, Russia  
Website: <http://i-us.ru/en>, e-mail: [i.us.spb@gmail.com](mailto:i.us.spb@gmail.com)  
Tel.: +7 - 812 494 70 02**INFORMATION PROCESSING AND CONTROL****Dumaev R. I., Molodyakov S. A., Utkin L. V.***3D medical image segmentation with persistent homology-based constraints*

2

**INFORMATION AND CONTROL SYSTEMS****Smirnov A. V., Levashova T. V., Teslya N. N.***Management of smart city road network configuration: a scenario based on collaboration patterns of decision-making participants*

13

**Tatarnikova T. M., Arkhiptsev E. D.***Hybrid method of time synchronization in distributed systems*

26

**SYSTEM AND PROCESS MODELING****Satsiuk A. V., Volodarets N. V.***Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system*

36

**HARDWARE AND SOFTWARE RESOURCES****Zhukov S. V., Kovaleva O. A., Kovalev S. V.***Multi-criteria analysis of web page optimization methods and their impact on search engine rankings*

45

**INFORMATION SECURITY****Bakeev R. N., Kuzmin V. N., Menisov A. B., Sabirov T. R.***Method for identifying software code vulnerabilities based on cluster analysis and contextual adaptation of large language models*

58

**INFORMATION ABOUT THE AUTHORS**

71

4(137)/2025

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

ИНФОРМАЦИОННО-  
УПРАВЛЯЮЩИЕ  
СИСТЕМЫ**Учредитель**

А. А. Востриков

**Издатель**Санкт-Петербургский государственный университет  
аэрокосмического приборостроения**Главный редактор**

Е. А. Крук,

д-р техн. наук, проф., Москва, РФ

**Ответственный секретарь**

О. В. Муравцова

**Редакционная коллегия:**

С. Д. Андреев,

д-р техн. наук, Тампере, Финляндия

В. Г. Анисимов,

д-р техн. наук, проф., Санкт-Петербург, РФ

Б. П. Безручко,

д-р физ.-мат. наук, проф., Саратов, РФ

Н. Блаунштейн,

д-р физ.-мат. наук, проф., Беэр-Шева, Израиль

М. В. Буэдалов,

канд. техн. наук, научный сотрудник, Санкт-Петербург, РФ

Л. С. Джайн,

д-р наук, проф., Канберра, Австралия

А. Н. Дудин,

д-р физ.-мат. наук, проф., Минск, Беларусь

И. И. Думер,

д-р наук, проф., Риверсайд, США

А. И. Зейфман,

д-р физ.-мат. наук, проф., Вологда, РФ

К. Кристоделу,

д-р наук, проф., Альбукерке, Нью-Мексико, США

А. А. Мюллари,

д-р наук, профессор, Гренада, Вест-Индия

К. Е. Самуйлов,

д-р техн. наук, проф., Москва, РФ

Д. Себерри,

д-р наук, проф., Волонгонг, Австралия

М. Б. Сергеев,

д-р техн. наук, проф., Санкт-Петербург, РФ

А. В. Смирнов,

д-р техн. наук, проф., Санкт-Петербург, РФ

Т. Суткинью,

д-р наук, доцент, Джокьякарта, Индонезия

А. А. Тюгашев,

д-р техн. наук, проф., Самара, РФ

М. Н. Фаворская,

д-р техн. наук, проф., Красноярск, РФ

Л. Фортуна,

д-р наук, проф., Катания, Италия

А. Л. Фрадков,

д-р техн. наук, проф., Санкт-Петербург, РФ

А. Е. Храмов,

д-р физ.-мат. наук, Иннополис, РФ

А. А. Шалыто,

д-р техн. наук, проф., Санкт-Петербург, РФ

А. П. Шепета,

д-р техн. наук, проф., Санкт-Петербург, РФ

Ю. И. Шокин,

акад. РАН, д-р физ.-мат. наук, проф., Новосибирск, РФ

З. М. Юлдашев,

д-р техн. наук, проф., Санкт-Петербург, РФ

**Редактор:** А. Г. Ларионова**Корректор:** Т. В. Звертановская**Дизайн:** М. Л. Черненко, Ю. В. Умницына**Компьютерная верстка:** Ю. В. Умницына**Адрес редакции:** 190000, г. Санкт-Петербург,

ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ

Тел.: (812) 494-70-02, эл. адрес: ius.spb@gmail.com,

сайт: http://i-us.ru

**ОБРАБОТКА ИНФОРМАЦИИ И УПРАВЛЕНИЕ***Dumaev R. I., Molodyakov S. A., Utkin L. V.**3D medical image segmentation with persistent homology-based constraints*

2

**ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ***Смирнов А. В., Левашова Т. В., Тесля Н. Н.**Управление конфигурацией улично-дорожной сети умного города: сценарий на основе паттернов коллективных действий участников принятия решений*

13

*Татарникова Т. М., Архипцев Е. Д.**Гибридный метод синхронизации времени в распределенных системах*

26

**МОДЕЛИРОВАНИЕ СИСТЕМ И ПРОЦЕССОВ***Сацюк А. В., Володарец Н. В.**Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением*

36

**ПРОГРАММНЫЕ И АППАРАТНЫЕ СРЕДСТВА***Жуков С. В., Ковалева О. А., Ковалев С. В.**Многокритериальный анализ методов оптимизации веб-страниц и их влияние на ранжирование в поисковых системах*

45

**ЗАЩИТА ИНФОРМАЦИИ***Бакеев Р. Н., Кузьмин В. Н., Менисов А. Б., Сабиров Т. Р.**Метод определения уязвимостей программного кода на основе кластерного анализа и контекстной адаптации больших языковых моделей*

58

**СВЕДЕНИЯ ОБ АВТОРАХ**

71

Журнал входит в БД Scopus и в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

Сдано в набор 03.07.25. Подписано в печать 27.08.25. Дата выхода в свет: 29.08.2025.

Формат 60×841/8. Гарнитура CentSchbkCyrill BT. Печать цифровая.

Усл. печ. л. 8,7. Уч.-изд. л. 11,9. Тираж 1000 экз (1-й завод 50 экз.). Заказ № 264.

Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.

190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А.

Отпечатано в редакционно-издательском центре ГУАП.

190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А.

Распространяется бесплатно.

Журнал зарегистрирован в Министерстве РФ по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г.

Перерегистрирован в Роскомнадзоре.

Свидетельство о регистрации ПИ № ФС77-82226 от 23 ноября 2021 г.

© А. А. Востриков, 2025



## 3D medical image segmentation with persistent homology-based constraints

R. I. Dumaev<sup>a</sup>, Post-Graduate Student, [orcid.org/0009-0002-3674-4952](https://orcid.org/0009-0002-3674-4952)

S. A. Molodyakov<sup>a</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0003-2191-9449](https://orcid.org/0000-0003-2191-9449), [samolodyakov@mail.ru](mailto:samolodyakov@mail.ru)

L. V. Utkin<sup>a</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0002-5637-1420](https://orcid.org/0000-0002-5637-1420)

<sup>a</sup>Peter the Great St. Petersburg Polytechnic University, 29, Politekhnikheskaia St., 195251, Saint-Petersburg, Russian Federation

**Introduction:** Medical image segmentation is a widely researched field where neural networks serve as a basis for many medical data analysis and visualization processes. Traditional approaches to training segmentation models often rely on voxel-wise loss functions, which are insufficient for many segmentation tasks, and do not consider the topological correctness of the segmentation. Consequently, masks learned by such models may be spatially inconsistent, resulting in unrealistic features such as spurious connected components or holes. **Purpose:** To develop a method for training models to improve the quality of medical 3D image segmentation by leveraging persistent homology and comparing persistence diagrams during training. **Results:** We propose a method for training 3D image segmentation models, including persistent homology-based constraints and a loss function that is used to regularize the shape and edge reconstruction process of the mask. We present a filtration function based on the distance to the centroid of a binary mask to refine the shape and edge of the predicted mask. The analysis of the results obtained during the experiments on lung computed tomography data for segmentation and target structure extraction tasks has shown the effectiveness of our approach. The proposed approach not only improves the accuracy at the voxel level, but also preserves essential morphological properties, which is extremely important for subsequent tasks, such as nodule volume estimation and clinical decision making. **Practical relevance:** The use of the presented approach makes it possible to improve the quality of lung nodule segmentation using 3D CT images.

**Keywords** – medical image segmentation, shape prediction, topological data analysis, persistent homology, lung nodule segmentation.

**For citation:** Dumaev R. I., Molodyakov S. A., Utkin L. V. 3D medical image segmentation with persistent homology-based constraints. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 2–12. doi:10.31799/1684-8853-2025-4-2-12, EDN: ZRIITM

### Introduction

In the analysis and visualization of clinical data workflows, accurate segmentation of medical images is a crucial step. For example, lung nodule segmentation plays a vital role in computer-aided diagnosis (CAD) systems for lung cancer, as it provides essential information such as nodule sizes, shapes, and other critical medical characteristics [1]. By allowing for accurate volume estimation, reliable segmentation is essential for effective nodule management, particularly when calculating volume doubling time [2, 3]. The large volume of data generated by CT imaging, which can produce hundreds of images of the lungs per second, poses a significant challenge to radiologists, who must individually examine thousands of slices in each scan for diagnostic purposes. This process is not only time-consuming and labor-intensive but also susceptible to errors [4]. Furthermore, manual segmentation is prone to observer variability, highlighting the need for an efficient and accurate automated lung nodule segmentation method to enhance the early detection rate of lung cancer.

The works [5, 6] show that there has been significant progress in the segmentation of 2D and 3D

medical images. This was influenced by the advent of deep learning models and convolutional neural networks, an successful example of which is the U-Net architecture [7], which generalizes features at different levels (from local at the voxel level to global at the anatomical level) and is successfully applied in various domains. At the same time, it can be noted that less attention is paid to the process of model optimization [8] and most works use loss functions common in segmentation problems: cross-entropy, Jaccard index and Dice coefficient (which consider voxels independently and are insensitive to errors such as false connected components or holes). As a result, such loss functions are insufficient for many segmentation problems and do not take into account the topological correctness of the segmentation, which can be learned during the optimization process.

In clinical practice, the characteristics of pulmonary nodules, such as their sizes, shapes, and specific features like calcification, lung lobe, and ground-glass shadows, exhibit significant variability across different levels. In addition, the morphologic features seen on CT images often resemble those of vascular tissues, which may make accurate

segmentation of pulmonary nodules difficult and lead to spatial inconsistencies.

If spatial features are not taken into account when optimizing a convolutional neural networks, it can result in segmentations that lack spatial coherence, including spurious connected components or holes. Such errors can appear nonsensical, but they are often confined to object boundaries and may not significantly impact the assessment of specific clinical features. Nevertheless, for many downstream applications, such as nodule volume estimation [2, 3], it is essential to accurately capture these spatial features in order to ensure reliable results.

Capturing spatial consistency and boundary consistency in segmentation masks requires abstract and adaptable information that can accommodate the diverse range of cases encountered in clinical settings, rather than relying on specific examples from training data. Spatial prior consistency offers a global measure of segmentation consistency. Such measure can be applied to normal and complex pathological regions of an image, unlike conventional losses that treat voxels independently. However, while quantifying conventional loss functions is straightforward, estimating spatial prior consistency presents significant complexity and non-trivial challenges. Furthermore, explicitly integrating this measure into convolutional neural networks is hampered by the opacity of such models. We present a topological descriptor module and topological constraints to address the limitations and challenges described above.

The contributions can be summarized as follows:

- Recently, there have been some papers that use topological data analysis (TDA) together with machine learning [9]. Using TDA, it is possible to analyze the shape of objects in different dimensions and use it as an additional source of features. We propose a loss function that uses these features to help the model learn the shape and boundaries of a mask in a segmentation task.

- We propose a centroid loss that encourages the alignment of centroids between the predicted mask and the ground truth to improve optimization process.

- To capture boundary information of shape we introduce filtration based on distance to the centroid of binary mask.

## Related works

### Medical images segmentation

Accurate segmentation of lung nodules is a crucial component of CAD systems, as it enables the precise delineation of a target nodule's boundary. The primary objective of lung nodule segmentation is to provide detailed information about the

nodule, including its diameter, size, and semantic characteristics important for diagnosis [10]. In [11], authors highlight the importance of these details for malignancy assessment of the nodule and treatment planning. The segmentation method allows for the extraction of precise boundaries of the nodule and overall size [12]. The main challenge of this task is that lung nodules have various shapes, sizes, and delicate features. Recently, deep learning has become the most popular method in this area. In [13], the authors propose a novel model called the Dual-branch Residual Network (DB-ResNet), which enhances segmentation accuracy by integrating multi-view and multi-scale feature extraction techniques. This model can handle the complexities associated with different types of lung nodules: the proposed method involves the use of features capture from three different slices (upper, middle, lower) and central intensity pooling layer, which focuses on extracting intensity features from the center voxel of the nodule.

Zhou et al. [14] proposed a cascaded multistage framework and split the process into several stages. In the localization stage 2D-based model performs initial rough localization of nodules on axial CT slices. In aggregation and segmentation stages, the candidate nodule selection algorithm refines the location of nodules and reduce redundancy among candidates and segment final masks.

Yang et al. [15] considered the problem of capturing the variability of clinical practice where different radiologists may provide different segmentations for the same nodule and designed an uncertainty-aware attention mechanism that leverages the consensus and disagreements among multiple annotations to improve segmentation performance. Their model aims to provide a more nuanced understanding of areas with varying levels of confidence.

The results of such models can be used to perform subsequent evaluation of the characteristics of the nodule [16]. Observations from previous studies suggest that current methods tend to focus primarily on improving segmentation accuracy without taking into account the spatial and boundary consistency of the predicted segmentation mask.

### Segmentation with topological data analysis

Various approaches leveraging TDA have been developed to ensure the topological accuracy of objects in images [17, 18]. In [19], the authors proposed method measures the difference between high-level structural information which include concepts such as connectivity or holes captured by pretrained network for delineating curvilinear structures such as cell membranes. TDA has also been successfully applied for tumor segmentation [20] and cortical plate segmentation [21] in 3D MRI scan, cancer prediction on whole slide histology images [22], for ultrasound

imaging analysis [23] and object segmentation in microscopy images [24]. The first type of application involves designing segmentation methods that constrain their topology. For instance, researchers in [25] introduced a topological loss function, calculated using TDA, to supplement the standard loss function used in training segmentation models. By incorporating prior topological knowledge, the segmentation process is restricted to produce results that closely resemble a predefined shape. This concept has been successfully applied to segmenting the myocardium, which is known to have a ring-like shape, and to segmenting the placenta, which is recognized as a single connected component without holes.

An alternative approach, presented in [21, 26], involves training a neural network to minimize the difference between the persistence diagrams of the images and their corresponding ground truth segmentations. Rather than relying on explicit prior knowledge to constrain the persistence diagram, the authors introduce a topology-preserving loss function. It encourages the segmentation output to preserve the same topological characteristics as the ground truth, specifically by aligning the Betti numbers (which represent the number of connected components and holes) between the predicted and actual segmentation masks.

As an another application of TDA, it can be utilized to characterize the topological structure of image components. For example, in [27, 28], the authors use persistent homology to identify image patches that contain tumors, based on the presence of an increased number of holes and connected components. This approach is grounded in the fact that infected tissues often exhibit irregular nuclei

shapes and sizes, which can be distinguished from healthy tissues using TDA. Finally, it is worth noting that TDA is a rapidly evolving field with ongoing research and development of new methods.

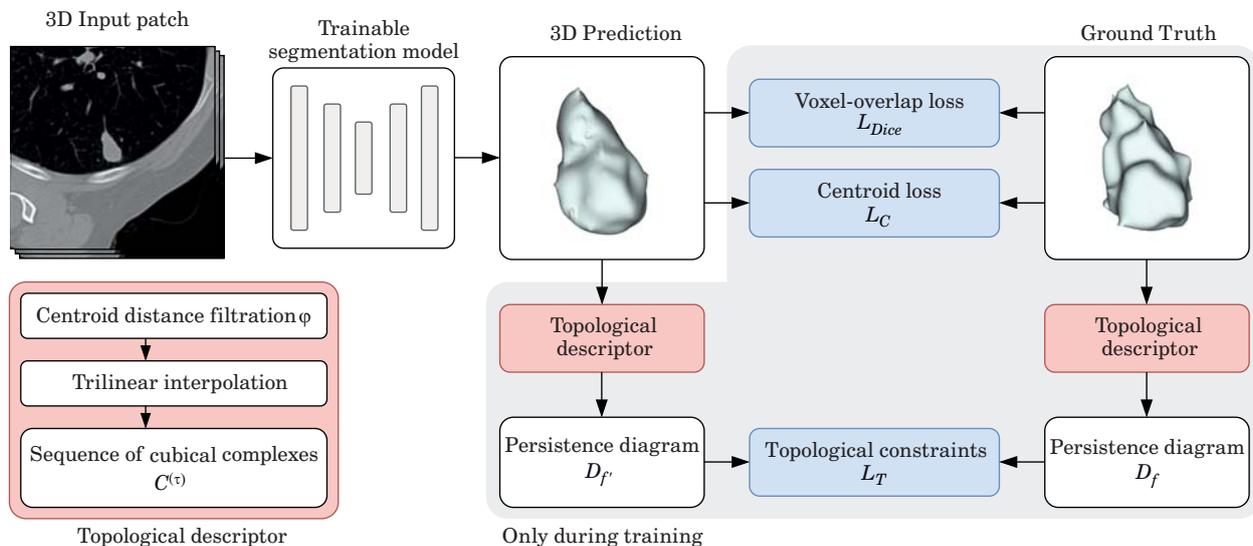
### Method for training 3D segmentation models

In Figure 1 we provide a flowchart of our proposed pipeline, which include the integration of TDA, particularly persistent homology, with a new filtration function based on distance to the centroid to refine shape and boundary topology. Given a predicted mask and 3D ground truth, we calculate topological features using topological descriptor with cubical persistent homology. We compare the resulting persistence diagrams using topological constraints and combine it with centroid loss and Dice loss.

The output of the trainable model and the ground truth segmentations are sequentially processed through a topological descriptor to efficiently capture the topology information of target instances. Then we compare the resulting persistence diagrams using topological constraints and combine it with centroid loss  $L_C$  and Dice loss  $L_{Dice}$ . It encourages the segmentation output of the trainable model to preserve the same topological characteristics as the ground truth.

### Topology assessment

Previous studies [29, 30] have shown that cubic complexes are effective for representing and exploring volumetric data. Topological features of different dimensions for cubical complex include



■ Fig. 1. Overview of the proposed method

connected components, cycles, and voids. We represent a volume  $V$  as a cubical complex  $C$ . A volume  $V$  can be described by a  $m$ -dimensional tensor with a shape of  $n_1 \times n_2 \times \dots \times n_m$ . Cubical complex represents an each voxel of a volume  $V$  as vertices and their spatial relationships (points, lines, squares and cubes).

The Betti number, represented as  $\beta_k$ , measures the number of  $k$ -dimensional topological features, such as connected components, loops, or voids, in a cubical complex  $V$ . Betti numbers only provide a count of features and do not account for their scale or how long they persist across different resolutions. Therefore, they are limited in applications requiring multi-scale analysis, although there are works that use them in image segmentation tasks as a topological prior [31, 32]. To address this limitation we utilize persistent diagrams to enable the calculation of topological features (capturing both their presence and their persistence over varying levels of detail) across multiple scales.

To analyze the structure of the volume, we apply a threshold  $\tau \in \mathbb{R}$  which filters the voxels based on their likelihood values. We get likelihood values using likelihood function  $f: V \rightarrow \mathbb{R}$ , learned by our model, evaluates each voxel  $x \in V$ , and assigns it a probability value indicating how likely it is to belong to a nodule. Specifically, a cubical complex  $C^{(\tau)}$  is formed by including all voxels  $x$  that satisfy  $f(x) \geq \tau$ . The topology of the resulting cubical complex changes only at a finite number of thresholds  $\tau \geq \dots \geq \tau_m$  (due to the volume is finite). Thus, by applying these thresholds, we obtain super-level set filtration – a sequence of nested cubic complexes:  $\emptyset \subseteq C^{(\tau_1)} \subseteq C^{(\tau_2)} \subseteq \dots \subseteq C^{(\tau_m)} = V$  which captures the changes of the volume’s topological features. In this way we can track pairs of thresholds at which  $k$ -dimensional features appear and disappear. We then define the persistence diagram  $D_f^{(k)}$  as the set of pairs of such thresholds:  $(\tau_i, \tau_j)$ , where  $\tau_i \geq \tau_j$  – creation and destruction of  $k$ -dimensional features with  $0 \leq k \leq m$ . By monitoring these features throughout the entire filtration, persistent homology offers a detailed and comprehensive characterization of the nodule’s topological structure.

In contrast to Betti numbers, which require a fixed threshold  $\tau$  for the likelihood function and

yield a single set of topological descriptors, persistence diagrams offer a more comprehensive representation of the data by encoding all possible thresholds simultaneously. This allows for the capture of additional geometrical details that would be lost with a single threshold choice. The persistence of a topological feature defined as  $pers(\tau_i, \tau_j) = |\tau_i - \tau_j|$  for a given tuple  $(\tau_i, \tau_j)$  in the persistence diagram, provides a measure of the feature’s scale or stability. Larger persistence values typically indicate more robust features that persist over a wider range of thresholds.

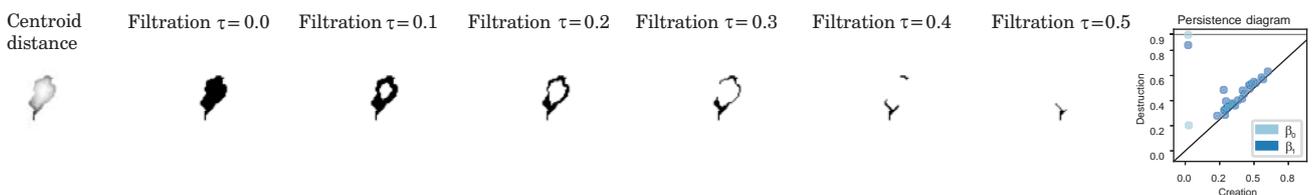
It is important to note that although the filtration can be computationally expensive, practical implementations have shown that persistent homology for cubical complexes can be computed efficiently, as highlighted in previous work [29].

We showcase an example of a filtration process from a centroid distance function in Fig. 2. Given a centroid distance function of the binary mask, as the threshold increases, various features appear and disappear. Each point on the persistence diagram represents a feature and has a creation and destruction value – interval in which it appears and disappears. Note that this is for illustrative purposes only, and in reality the resulted persistence diagram may look different, since cubical complex for this example is built from only one slice of the 3D nodule patch.

### Topological descriptor

As previously noted, pixel-wise loss is insensitive to capturing spatial consistency and boundary consistency in segmentation masks. To explicitly incorporate such priors into parameter optimization and quantitatively represent them, we present a topological descriptor module.

At the input, the module receives a binary mask  $V$  and transforms it using the filter function  $\phi$ . Then, for computational efficiency, we downsample each sample to reduce the dimensionality using bilinear and trilinear interpolation for 2D and 3D respectively. We provide an additional information about the impact of interpolation step in Experiments section. Based on the interpolated result, we calculate a cubic complex  $C$ . This module is used to obtain topological features, which are then compared with topological constraints.



■ Fig. 2. An example of the filtration process

The choice of filtration in persistent homology determines the type of information that is captured about cycles, ranging from purely topological to a combination of topological and geometric information. Specific topological invariants exhibit robustness under affine transformations, whereas geometric invariants do not necessarily possess this property. As a result, the choice of filtration has significant implications for the persistence and importance of cycles, influencing which cycles are deemed significant and which are considered noise and less relevant [33]. To capture boundary information of shape we introduce filtration based on distance to the centroid of binary mask.

Lets  $V$  be a gray-scale volume,  $V = [x_{dhw}]$  where  $x_{dhw}$  is the gray-scale value of the voxel  $(d, h, w)$ ,  $d \in \{1, 2, \dots, n_1\}$ ,  $h \in \{1, 2, \dots, n_2\}$ ,  $w \in \{1, 2, \dots, n_3\}$ . First, we define a radial filtration function as  $l_2$ -norm distance between centroid of the binary mask and the reference point:

$$\varphi_r = \begin{cases} \|(d, h, w) - (d_0, h_0, w_0)\|_2, & x_{dhw} \geq x_0; \\ 0, & \text{otherwise,} \end{cases}$$

where  $(d, h, w)$  is the reference point for which the distance is calculated,  $x_0$  is the gray-scale threshold, we take threshold value  $x_0 = 0.5\max(V)$ . We use centroid of the binary mask as the reference point

$(d_0, h_0, w_0)$  for radial filtration. The centroid is defined as a simple spatial average.

To capture information not only about topological features, but also about shape information the final filtration function defined as  $\varphi = \varphi_r(1 - \varphi_{edt})$ , where  $\varphi_{edt}$  is euclidean distance transform filtration which calculates the distance by replacing each foreground element on the mask with its shortest distance to the background.

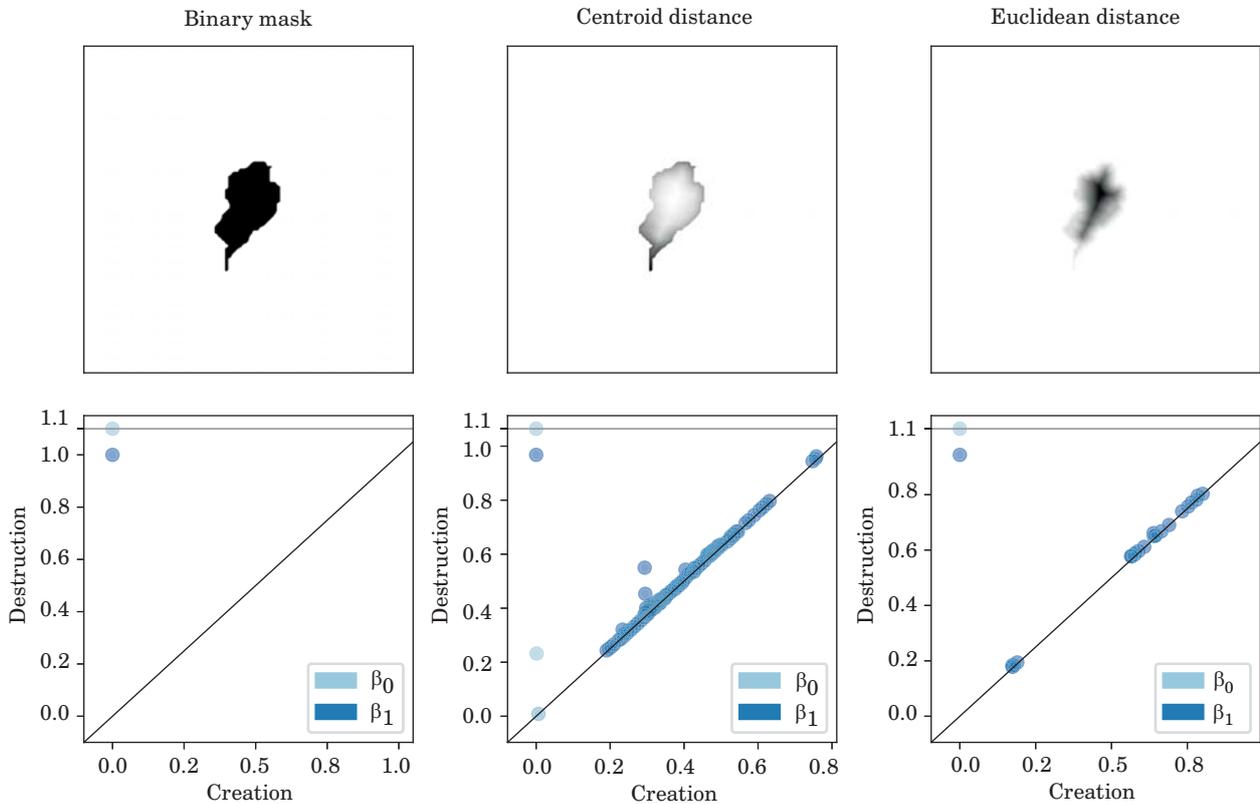
In Figure 3 we visualize different filtration functions for the same nodule slice and their persistence diagrams. The first plot shows a slice of the nodule patch, the next two plots respectively show filtrations by centroid distance and euclidean distance for this slice, and three plots of persistence diagrams.

### Topological constraints

Given a true likelihood function  $f$  and a predicted likelihood function  $f'$  and corresponding persistence diagrams  $D$  and  $D'$  we use Wasserstein distance to compare these diagrams. So, our topological loss can be defined as

$$L_T = \sum_{k=0}^d W_p(D_f^{(k)}, D_{f'}^{(k)}),$$

where  $W_p(D_f^{(i)}, D_{f'}^{(i)})$  is a Wasserstein distance between persistence diagrams. This loss term



■ Fig. 3. Comparison of different filtrations and their persistence diagrams

encourages the model to reduce the distance between  $f$  and  $f'$  with respect to their topological shape and boundary information.

In the training pipeline with topological constraints during optimization of the  $L_T$  the Dice loss  $L_{Dice}$  landscape is flat in regions where there is no overlap between the prior and the ground truth due to, we regularize the shape and boundary reconstruction. The result is a limited and local modification of the prior that does not extend to the point of intersecting with the ground truth, and as a result the loss remains unchanged. To prevent the loss landscape from entering flat regions we propose a centroid loss that encourages the alignment of centroids between the predicted mask and the ground truth.

Proposed centroid loss can be defined as

$$L_C = \sqrt{\|(d_f, h_f, w_f) - (d_{f'}, h_{f'}, w_{f'})\|_2},$$

where  $(d_f, h_f, w_f)$  and  $(d_{f'}, h_{f'}, w_{f'})$  are centroids of the prior and the ground truth respectively.

To jointly optimize the segmentation network the final loss can be defined as

$$L = L_{Dice} + \lambda_T L_T + \lambda_C L_C,$$

where  $\lambda_T$  and  $\lambda_C$  controls the impact of the corresponding loss terms.

## Experiments

### Datasets

The performance of our proposed training pipeline was evaluated using the publicly available LIDC-IDRI dataset [34], a comprehensive collection of CT scans from 1,010 patients, featuring annotated nodules that were identified and outlined by multiple radiologists from various medical institutions across the United States, employing a range of imaging protocols and scanner models from four different manufacturers.

We selected nodules that had a diameter in a range 5–50 mm, had minimum 1000 voxels in annotation, correct annotations and slice spacing in related scan and were identified by at least one radiologist. As a result, we had 393 nodule samples from LIDC-IDRI dataset, which divided into a training dataset of 282 samples, validation dataset of 32 samples and test dataset of 79 samples. We make sure that no sample is included in either the training, validation or test dataset.

### Evaluation metrics

We use three main instance-level evaluation metrics to measure the instance segmentation performance of the comparison models: Jaccard index

(JI), Dice score (DSC) and Hausdorff distance (HD) [35]. Jaccard index can be calculated as

$$JI = \frac{|Y_{true} \cap Y_{pred}|}{|Y_{true} \cup Y_{pred}|}.$$

It measures the ratio of the intersection and the union between the ground truth mask and predicted mask. Dice score focuses on the ratio of the intersection and the sum of the sets and penalizes differences less strictly because the denominator doesn't grow as fast as in the Jaccard index. Dice score can be defined as

$$DSC = \frac{2|Y_{true} \cap Y_{pred}|}{|Y_{true}| + |Y_{pred}|}.$$

Also we obtained additional performance metrics based on Minkowski functionals: volume error, surface area error, mean curvature error, which are calculated as the difference between the values of the corresponding  $n$ -th functional between the predicted mask and the ground truth:

$$M_n^{error} = \frac{|M_n^f - M_n^{f'}|}{M_n^f},$$

where  $M_n^f$  and  $M_n^{f'}$  are corresponding Minkowski functional of the ground truth and the prior.

### Implementation details

To extract the volume of interest for each nodule in our image processing workflow, we employ the pylidc package, a specialized library designed for the LIDC-IDRI dataset [36]. All patches are resampled to  $0.75 \times 0.75 \times 0.75$  mm physical space and cropped to the resolution  $96 \times 96 \times 96$  voxels using linear interpolation. Then we clamp intensity values (Hounsfield units) of volumes into a range  $[-1000, 1000]$  and rescale intensity values to a range  $[0, 1]$  with contrast stretching, as done in previous studies [37]. To implement processing pipeline for 3D data we use TorchIO library [38]. To calculate Minkowski functionals we use QuantImPy library [39].

The DeepLabV3 [40] architecture is employed as our target model, and in order to guarantee a fair comparison, we maintain a consistent implementation and hyperparameters setting throughout the entirety of our experiments, we employ "Weights and Biases" for tracking experiments [41]. The network are trained from imagenet weights for 200 epochs with a batch size of 28. We use the Adam optimizer with a dynamic learning rate (cosine annealing warm restarts [42]) from the start value  $6 \times 10^{-5}$  to the maximum value  $1 \times 10^{-4}$ . The op-

optimizer has the first moment estimate 0.9 and the second moment estimate 0.999.

Also we optimize  $\lambda_T$  and  $\lambda_C$  parameters on an independent data set, as a result we used  $\lambda_T = 0.02$  and  $\lambda_C = 1.0$  for all experiments. These parameter values are chosen because the topological loss term is based on the Wasserstein distance between persistence diagrams and compares multi-scale topological features across the entire volume, which can yield relatively large numerical values. Larger  $\lambda_T$  parameter values can destabilize the model training process (topological features can be noisy at early training stages). Target shape for the interpolation step in topological descriptor was  $18 \times 18 \times 18$  voxels because it's optimal in terms of performance and errors.

We augment the training data using standard image data augmentation techniques. This included randomly applying transformations such as reverse the order of elements in a volume along the axes, random rotating the input patch in a range  $[-25, 25]$  degrees and random affine transformation. To maintain consistency, we apply identical transformations to all slices within a nodule volume. Furthermore, we randomize the order of training examples in the each epoch. Our models and training pipeline are built using PyTorch and data processing pipeline is built with TorchIO and Torchvision libraries. The source code for the implementation of models and the training pipeline, data processing and dataset preparation is available at the link: [https://github.com/dumaevrinat/ph\\_constraints](https://github.com/dumaevrinat/ph_constraints).

### Experiment results and discussion

To evaluate the benefits of our topological constraints, we perform the same experiment twice: in first run we use Dice loss only, and then added  $\lambda_T L_T$  and  $\lambda_C L_C$  loss terms. As shown in Table, our training pipeline with topological constraints outperforms the baseline across all three main metrics. We also provide error comparison based on Minkowski functionals. The volume error, surface area error, and mean curvature error are all also slightly lower for the approach with topological constraints. We perform a paired Wilcoxon signed-rank test between the two error distributions, the obtained p-values are also presented in the Table. We show a detailed comparison of  $1 - DSC$  (fig. 4, a) and  $1 - JI$  (fig. 4, b) evaluation metrics on different groups of nodules divided into quartiles by volume and surface area.

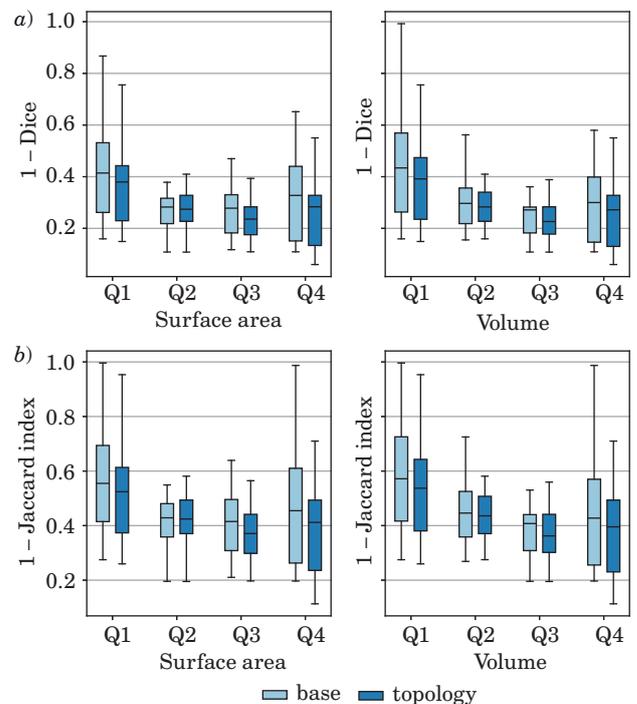
The results of comparing two approaches to model training, with and without topological constraints, indicate that incorporating topological constraints leads to improved lung nodule segmentation performance.

The effectiveness of the proposed segmentation method with topological constraints is illustrated in

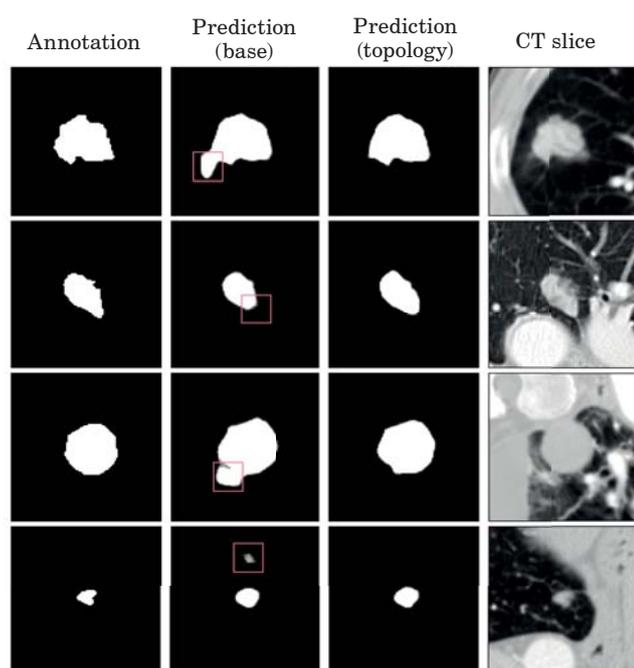
the Fig. 5. The segmentation results (one slice from the whole segmentation volume) for four example nodules from the test dataset are compared against the baseline approach, with invalid regions shown as red rectangles. The baseline method exhibits inconsistencies such as fragmented boundaries and isolated false-positive regions. In contrast, method based on persistent homology-based constraints produces more closely segmentations to the ground truth. This improvement is especially noticeable in the absence of topological errors such as holes or unconnected components.

■ Lung nodule segmentation performance of different training pipelines

Evaluation metric	Without topological constraints	With topological constraints	p-value
$1 - DSC$	$0.33 \pm 0.20$	$0.29 \pm 0.16$	0.0103
$1 - JI$	$0.46 \pm 0.19$	$0.43 \pm 0.17$	0.0056
HD	$7.74 \pm 5.05$	$6.09 \pm 5.70$	0.0094
$M_0$ (volume error)	$0.29 \pm 0.32$	$0.30 \pm 0.29$	0.0465
$M_1$ (surface area error)	$0.22 \pm 0.25$	$0.19 \pm 0.23$	0.0158
$M_2$ (mean curvature error)	$0.26 \pm 0.21$	$0.21 \pm 0.22$	0.0002



■ Fig. 4. Comparison of Dice (a) and Jaccard (b) loss metric for base and topological pipelines for samples divided into quartiles by surface area and volume



■ **Fig. 5.** Comparison of lung nodule segmentation results

The proposed method integrates persistent homology-based topological constraints into 3D medical image segmentation, demonstrating improved performance over traditional pixel-wise loss functions. By leveraging TDA, our approach ensures that predicted segmentations adhere to expected morphological properties and boundary consistency. Below, we discuss our results, compare our method with existing approaches, and outline its limitations and potential applications.

Most deep learning-based segmentation methods focus on improving accuracy through architectural modifications or post-processing. These approaches often ignore topological correctness. Recent works have explored topology-aware losses, but many of them rely on simplifying assumptions such as fixed Betti numbers [25], which lack the multi-scale flexibility of persistent homology. In contrast, our method captures topological and shape representation features across varying scales through persistence diagrams.

Despite the above advantages, our method has several limitations. The topological loss assumes

high-quality ground truth annotations. In practice, variability in medical annotations from multiple experts may require robustness to noisy labels. Calculating persistent homology for high-resolution 3D volumes is computationally intensive and requires CPU resources. Although downsampling mitigates this to some extent, future work may focus on exploring this limitation.

Although the effectiveness of the method has been confirmed in pulmonary lung nodules, further studies are needed for other types of lesions.

## Conclusion

In this paper, we proposed persistent homology-based topological constraints and their integration into medical image segmentation represents a significant advancement in the field, particularly for applications such as lung nodule detection. By employing a new filtration function based on centroid distance, the proposed method effectively refines shape and boundary representation and improves the accuracy of segmentation.

Empirical results (see Table) for lung nodule segmentation using 3D computed tomography images validate the effectiveness of our framework. The results indicate that this approach not only improves pixel-level accuracy but also preserves essential morphological properties, which are critical for downstream applications such as volume estimation and clinical decision-making.

Future studies and improvements in topology-aware methods may further increase the accuracy and usefulness of automated segmentation tools, ultimately leading to improved patient outcomes in clinical settings. The current evolution of TDA within machine learning has great potential to improve understanding and interpretation of complex medical images.

## Financial support

The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation as part of state assignments, topic FSEG-2024-0027.

## References

1. Wang S., Zhou M., Liu Z., Liu Z., Gu D., Zang Y., Dong D., Gevaert O., Tian J. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical Image Analysis*, 2017, vol. 40, pp. 172–183. doi:10.1016/j.media.2017.06.014
2. de Margerie-Mellon C., Chassagnon G. Artificial intelligence: A critical review of applications for lung nodule and lung cancer. *Diagnostic and Interventional Imaging*, 2023, vol. 104, no. 1, pp. 11–17. doi:10.1016/j.diii.2022.11.001
3. Callister M., Baldwin D., Akram A., Barnard S., Cane P., Draffan J., Franks K., Gleeson F., Graham R., Malhotra P. British thoracic society guidelines for the

- investigation and management of pulmonary nodules: Accredited by nice. *Thorax*, 2015, vol. 70, no. Suppl 2, pp. ii1–ii54. doi:10.1136/thoraxjnl-2015-207168
4. Dong T., Wei L., Nie S. Research progress of lung nodule segmentation based on CT images. *Journal of Image and Graphics*, 2021. Available at: <https://api.semanticscholar.org/CorpusID:273256366> (accessed 15 December 2024).
  5. Agnes S. A., Anitha J. Efficient multiscale fully convolutional unet model for segmentation of 3D lung nodule from CT image. *Journal of Medical Imaging*, 2022, vol. 9, no. 5, pp. 052402–052402. doi:10.1117/1.JMI.9.5.052402
  6. Chartrand G., Cheng P. M., Vorontsov E., Drozdal M., Turcotte S., Pal C. J., Kadoury S., Tang A. Deep learning: A primer for radiologists. *Radiographics*, 2017, vol. 37, no. 7, pp. 2113–2131. doi:10.1148/rg.2017170077
  7. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *18th International Conference Medical Image Computing and Computer-Assisted Intervention – MICCAI*, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4\_28
  8. Duan J., Bello G., Schlemper J., Bai W., Dawes T. J., Biffi C., de Marvao A., Doumoud G., O'Regan D. P., Rueckert D. Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *IEEE Transactions on Medical Imaging*, 2019, vol. 38, no. 9, pp. 2151–2164. doi:10.1109/TMI.2019.2905196
  9. Hensel F., Moor M., Rieck B. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 2021, vol. 4, Article 681108.
  10. Xie H., Yang D., Sun N., Chen Z., Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition*, 2019, vol. 85, pp. 109–119. doi:10.1016/j.patcog.2018.08.006
  11. Pezzano G., Ripoll V. R., Radeva P. Cole-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation. *Computer Methods and Programs in Biomedicine*, 2021, vol. 198, Article 105792. doi:10.1016/j.cmpb.2020.105792
  12. Wu D., Lu L., Bi J., Shinagawa Y., Boyer K., Krishnan A., Salganicoff M. Stratified learning of local anatomical context for lung nodules in CT images. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2791–2798. doi:10.1109/CVPR.2010.5540027
  13. Cao H., Liu H., Song E., Hung C.-C., Ma G., Xu X., Jin R., Lu J. Dual-branch residual network for lung nodule segmentation. *Applied Soft Computing*, 2020, vol. 86, Article 105934. doi:10.1016/j.asoc.2019.105934
  14. Zhou Z., Gou F., Tan Y., Wu J. A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries. *IEEE Journal of Biomedical and Health Informatics*, 2022, vol. 26, no. 11, pp. 5619–5630. doi:10.1109/JBHI.2022.3196842
  15. Yang H., Wang Q., Zhang Y., An Z., Liu C., Zhang X., Zhou S. K. Lung nodule segmentation and uncertain region prediction with an uncertainty-aware attention mechanism. *IEEE Transactions on Medical Imaging*, 2023, vol. 43, no. 4, pp. 1284–1295. doi:10.1109/TMI.2023.3348698
  16. Dumaev R. I., Molodyakov S. A., Utkin L. V. A model for explainable malignancy assessment of pulmonary nodules on CT images. *Artificial Intelligence and Decision Making*, 2024, no. 4, pp. 123–134.
  17. Byrne N., Clough J. R., Valverde I., Montana G., King A. P. A persistent homology-based topological loss for CNN-based multiclass segmentation of CMR. *IEEE Transactions on Medical Imaging*, 2022, vol. 42, no. 1, pp. 3–14. doi:10.1109/TMI.2022.3207234
  18. Hu X., Li F., Samaras D., Chen C. Topology-preserving deep image segmentation. *Advances in Neural Information Processing Systems*, 2019, vol. 32, pp. 5657–5668.
  19. Mosinska A., Marquez-Neila P., Kozin'ski M., Fua P. Beyond the pixel-wise loss for topology-aware delineation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3136–3145. doi:10.1109/CVPR.2018.00331
  20. Demir A., Massaad E., Kiziltan B. Topology-aware focal loss for 3D image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 580–589. doi:10.1109/CVPR52729.2023.00061
  21. de Dumast P., Kebiri H., Dunet V., Koob M., Cuadra M. B. Multi-dimensional topological loss for cortical plate segmentation in fetal brain MRI. *arXiv preprint arXiv:2208.07566*, 2022.
  22. Skaf Y., Laubenbacher R. Topological data analysis in biomedicine: A review. *Journal of Biomedical Informatics*, 2022, vol. 130, Article 104082. doi:10.1016/j.jbi.2022.104082
  23. Singh Y., Farrelly C. M., Hathaway Q. A., Leiner T., Jagtap J., Carlsson G. E., Erickson B. J. Topological data analysis in medical imaging: Current state of the art. *Insights into Imaging*, 2023, vol. 14, no. 1, p. 58. doi:10.1186/s13244-023-01410-z
  24. Wen B., Zhang H., Bartsch D.-U. G., Freeman W. R., Nguyen T. Q., An C. Topology-preserving image segmentation with spatial-aware persistent feature matching. *arXiv preprint arXiv:2412.02076*, 2024.
  25. Clough J. R., Byrne N., Oksuz I., Zimmer V. A., Schnabel J. A., King A. P. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, vol. 44, no. 12, pp. 8766–8778. doi:10.1109/TPAMI.2020.2994504
  26. Sun S., Wang Y., Yang J., Feng Y., Tang L., Liu S., Ning H. Topology-sensitive weighting model for myocardial segmentation. *Computers in Biology and Med-*

- icine*, 2023, vol. 165, Article 107286. doi:10.1016/j.combiomed.2023.107286
27. Qaiser T., Tsang Y.-W., Taniyama D., Sakamoto N., Nakane K., Epstein D., Rajpoot N. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical Image Analysis*, 2019, vol. 55, pp. 1–14. doi:10.1016/j.media.2019.04.005
  28. Qaiser T., Sirinukunwattana K., Nakane K., Tsang Y.-W., Epstein D., Rajpoot N. Persistent homology for fast tumor segmentation in whole slide histology images. *Procedia Computer Science*, 2016, vol. 90, pp. 119–124. doi:10.1016/j.procs.2016.07.040
  29. François A., Tinarrage R. Train-free segmentation in MRI with cubical persistent homology. *arXiv preprint arXiv:2401.01160*, 2024.
  30. Rieck B., Yates T., Bock C., Borgwardt K., Wolf G., Turk-Browne N., Krishnaswamy S. Uncovering the topology of time-varying fMRI data using cubical persistence. *Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 6900–6912.
  31. Stucki N., Paetzold J. C., Shit S., Menze B., Bauer U. Topologically faithful image segmentation via induced matching of persistence barcodes. *International Conference on Machine Learning, PMLR*, 2023, pp. 32698–32727.
  32. Stucki N., Bürgin V., Paetzold J. C., Bauer U. Efficient Betti matching enables topology-aware 3D segmentation via persistent homology. *arXiv preprint arXiv:2407.04683*, 2024.
  33. Turkes R., Nys J., Verdonck T., Latre S. Noise robustness of persistent homology on greyscale images, across filtrations and signatures. *PLOS ONE*, 2021, vol. 16, no. 9, Article e0257215. doi:10.1371/journal.pone.0257215
  34. Armato S. G. III, McLennan G., Bidaut L., McNitt-Gray M. F., Meyer C. R., Reeves A. P., Zhao B., Aberle D. R., Henschke C. I., Hoffman E. A., Kazerooni E. A., MacMahon H., Van Beek E. J. R., Yankelevitz D., Biancardi A. M., Bland P. H., Brown M. S., Engelmann R. M., Laderach G. E., Clarke L. P. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 2011, vol. 38, no. 2, pp. 915–931. doi:10.1118/1.3528204
  35. Karimi D., Salcudean S. E. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 2019, vol. 38, no. 2, pp. 499–513.
  36. Hancock M. C., Magnan J. F. Lung nodule malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms: Probing the lung image database consortium dataset with two statistical learning methods. *Journal of Medical Imaging*, 2016, vol. 3, no. 4, Article 044504. doi:10.1117/1.JMI.3.4.044504
  37. Isensee F., Petersen J., Klein A., Zimmerer D., Jaeger P. F., Kohl S., Wasserthal J., Koehler G., Norajitra T., Wirkert S. nnU-Net: Self-adapting framework for U-Net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.
  38. Pérez-García F., Sparks R., Ourselin S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 2021, Article 106236. Available at: <https://www.sciencedirect.com/science/article/pii/S0169260721003102> (accessed 15 December 2024).
  39. Boelens A. M., Tchelepi H. A. Quantimpy: Minkowski functionals and functions with Python. *SoftwareX*, 2021, vol. 16, Article 100823. doi:10.1016/j.softx.2021.100823
  40. Chen L.-C. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
  41. Biewald L. *Experiment tracking with weights and biases*. 2020. Software available from wandb.com. Available at: <https://www.wandb.com/> (accessed 15 December 2024).
  42. Loshchilov I., Hutter F. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

УДК 004.852

doi:10.31799/1684-8853-2025-4-2-12

EDN: ZRIITM

### Сегментация медицинских 3D-изображений с использованием персистентных гомологий

Р. И. Думаев<sup>а</sup>, аспирант, orcid.org/0009-0002-3674-4952

С. А. Молодяков<sup>а</sup>, доктор техн. наук, профессор, orcid.org/0000-0003-2191-9449, samolodyakov@mail.ru

Л. В. Уткин<sup>а</sup>, доктор техн. наук, профессор, orcid.org/0000-0002-5637-1420

<sup>а</sup>Санкт-Петербургский политехнический университет Петра Великого, Политехническая ул., 29, Санкт-Петербург, 195251, РФ

**Введение:** традиционные подходы к обучению моделей сегментации часто опираются на функции потерь по вокселям, что недостаточно для многих задач сегментации, и не учитывают топологическую корректность сегментации. Следовательно, маски, полученные с помощью таких моделей, могут быть пространственно несогласованными, что приводит к появлению нереалистичных харак-

теристик, таких как ложные связанные компоненты или дыры. **Цель:** разработать метод обучения моделей для повышения качества сегментации медицинских 3D-изображений за счет использования персистентных гомологий и сравнения диаграмм персистентности в процессе обучения. **Результаты:** предложен метод обучения моделей сегментации 3D-изображений, включающий ограничения на основе персистентных гомологий и функцию потерь, которая применяется для регуляризации процесса восстановления формы и краев маски. Представлена функция фильтрации, основанная на расстоянии до центроида бинарной маски, для уточнения формы и границы предсказанной маски. Анализ результатов, полученных в ходе экспериментов на данных компьютерной томографии легких для задачи сегментации и выделения целевых структур, показал эффективность предложенного метода. Он не только повышает точность на уровне вокселей, но и сохраняет существенные морфологические свойства, что крайне важно для последующих задач, таких как оценка объема узелков и принятие клинических решений. **Практическая значимость:** использование представленного метода позволило повысить качество сегментации легочных узелков по 3D-изображениям компьютерной томографии.

**Ключевые слова** – сегментация медицинских изображений, предсказание формы, топологический анализ данных, персистентная гомология, сегментация узелков в легких.

**Для цитирования:** Dumaev R. I., Molodyakov S. A., Utkin L. V. 3D medical image segmentation with persistent homology-based constraints. *Информационно-управляющие системы*, 2025, № 4, с. 2–12. doi:10.31799/1684-8853-2025-4-2-12, EDN: ZRIITM

**For citation:** Dumaev R. I., Molodyakov S. A., Utkin L. V. 3D medical image segmentation with persistent homology-based constraints. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 2–12. doi:10.31799/1684-8853-2025-4-2-12, EDN: ZRIITM

---

---

### УВАЖАЕМЫЕ АВТОРЫ!

Научные базы данных, включая Scopus и Web of Science, обрабатывают данные автоматически. С одной стороны, это ускоряет процесс обработки данных, с другой – различия в транслитерации ФИО, неточные данные о месте работы, области научного знания и т. д. приводят к тому, что в базах оказывается несколько авторских страниц для одного и того же человека. В результате для всех по отдельности считаются индексы цитирования, что снижает рейтинг ученого.

Для идентификации авторов в сетях Thomson Reuters проводит регистрацию с присвоением уникального индекса (ID) для каждого из авторов научных публикаций.

Процедура получения ID бесплатна и очень проста, есть возможность провести регистрацию на 12 языках, включая русский (чтобы выбрать язык, кликните на зеленое поле сверху справа на стартовой странице): <https://orcid.org>

---

---



## Управление конфигурацией улично-дорожной сети умного города: сценарий на основе паттернов коллективных действий участников принятия решений

А. В. Смирнов<sup>а</sup>, доктор техн. наук, профессор, [orcid.org/0000-0001-8364-073X](https://orcid.org/0000-0001-8364-073X)

Т. В. Левашова<sup>а</sup>, канд. техн. наук, старший научный сотрудник, [orcid.org/0000-0002-1962-7044](https://orcid.org/0000-0002-1962-7044), [tatiana.levashova@iias.spb.su](mailto:tatiana.levashova@iias.spb.su)

Н. Н. Тесля<sup>а</sup>, канд. техн. наук, старший научный сотрудник, [orcid.org/0000-0003-0619-8620](https://orcid.org/0000-0003-0619-8620)

<sup>а</sup>Санкт-Петербургский Федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 191718, РФ

**Введение:** в процессе совместной деятельности участников цифровых сообществ, состоящих из людей и программных агентов, часто возникают повторяющиеся задачи, для решения которых могут быть использованы типовые модели поведения участников (паттерны коллективных действий). **Цель:** разработать модели эффективной организации целенаправленных действий участников цифровых сообществ и сценарий использования моделей при формировании рекомендаций по управлению конфигурацией улично-дорожной сети умного города как разновидности социотехнической системы. **Результаты:** разработана концептуальная модель паттерна коллективных действий, упрощающая обработку контекстной информации и взаимодействие участников цифрового сообщества за счет однородного представления информации, используемой в спецификациях паттернов различных видов. Разработана обобщенная модель выполнения задачи принятия решений на основе паттернов коллективных действий, позволяющая организовать выбор паттернов в процессе коллективного выполнения задачи. Предложен сценарий решения участниками цифрового сообщества задачи формирования рекомендаций по управлению конфигурацией улично-дорожной сети умного города, подтверждающий адекватность разработанных моделей. **Практическая значимость:** результаты исследований вносят вклад в решение проблемы управления конфигурацией социотехнических систем. Они предоставляют модели для эффективной организации целенаправленных коллективных действий участников цифровых сообществ в процессе совместного решения ими задачи управления конфигурацией как задачи принятия решений и могут быть использованы, например, для формирования рекомендаций по управлению конфигурациями таких систем, как умный город, аэропорт и т. п.

**Ключевые слова** – умный город, социотехническая система, цифровое сообщество, улично-дорожная сеть, поддержка принятия решений, паттерны коллективных действий.

**Для цитирования:** Смирнов А. В., Левашова Т. В., Тесля Н. Н. Управление конфигурацией улично-дорожной сети умного города: сценарий на основе паттернов коллективных действий участников принятия решений. *Информационно-управляющие системы*, 2025, № 4, с. 13–25. doi:10.31799/1684-8853-2025-4-13-25, EDN: ZOZUHI

**For citation:** Smirnov A. V., Levashova T. V., Teslya N. N. Management of smart city road network configuration: a scenario based on collaboration patterns of decision-making participants. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 13–25 (In Russian). doi:10.31799/1684-8853-2025-4-13-25, EDN: ZOZUHI

### Введение

Умные города представляют собой сложные социотехнические системы (СТС), способствующие повышению качества жизни за счет применения передовых технологий для принятия решений в сфере экономического и экологического использования городских систем жизнедеятельности [1]. СТС – это самоорганизующийся элемент гибридной техногенной среды, возникающий и развивающийся в результате взаимодействия и коммуникации человека, технической инфраструктуры и технологий. Помимо умных городов, примерами СТС являются аэропорты, энергетические, транспортные и другие системы, содержащие и реализующие компьютеризированные сетевые формы циклической комму-

никации, в которых осуществляется групповая деятельность участников коллектива [2].

Чтобы отвечать постоянным технологическим изменениям, характерным для современной эпохи, СТС должны быть способны развиваться в соответствии с этими изменениями. Преобразование существующей конфигурации СТС в новую является одной из функций управления конфигурациями. В данной статье под управлением конфигурацией понимается модификация конфигурации подсистем умных городов в целях ее адаптации к изменениям в технологической сфере и другим условиям, которые накладывает текущая ситуация. Концепция умного города предполагает повышение уровня комфорта горожан, часто достигаемое благодаря усовершенствованиям, которые учитывают

последние технологические изменения [3, 4], что подтверждает актуальность предлагаемых в данной работе исследований.

Одним из направлений развития общества, возникших благодаря прогрессу информационных технологий, являются цифровые сообщества. Эти сообщества, сложившиеся вследствие появления и интенсивного развития социальных сетей и объединяющие в себе географически разрозненных, но связанных общими интересами участников [5], приобретают всю большую значимость. Такие сообщества представляют собой сетевые системы, в которых цифровые технологии поддерживают взаимодействие людей и организаций, объединившихся в целях совместного решения проблем [5, 6]. Применительно к концепции умного города эти сообщества могут рассматриваться как коллективы, участники которых осуществляют совместную деятельность (коллективные действия) по управлению конфигурациями соответствующих подсистем.

Одним из аспектов концепции умного города является городское планирование, включающее в себя, помимо прочих, задачу управления конфигурацией улично-дорожной сети (УДС) [7]. На решение этой задачи участниками цифрового сообщества направлена данная статья. Применительно к рассматриваемой задаче цифровое сообщество реализует функции рабочей группы комиссии по обеспечению безопасности дорожного движения при администрации города (района). Эта комиссия является координационным органом, образованным для обеспечения согласованных действий исполнительных органов государственной власти различных уровней, ведомств, предприятий, учреждений, организаций независимо от форм собственности и общественных объединений в области обеспечения безопасности дорожного движения. Рабочая группа включает в себя специалистов (представителей) комитетов, управлений, инспекций и других административных образований, руководители которых входят в состав комиссии. Данная группа занимается формированием рекомендаций по обеспечению безопасности дорожного движения.

Цифровое сообщество рассматривает задачу управления конфигурацией УДС умного города в части адаптации технических средств организации дорожного движения к новым условиям как задачу принятия решений на основе паттернов коллективных действий. Решения, поддерживающие деятельность умного города, должны быстро приниматься и при этом быть эффективными [8, 9], что мотивировало обращение к таким паттернам. В процессе совместной деятельности часто возникают повторяющиеся однотипные проблемы, а паттерны коллективных действий предлагают готовые ответы для подобных про-

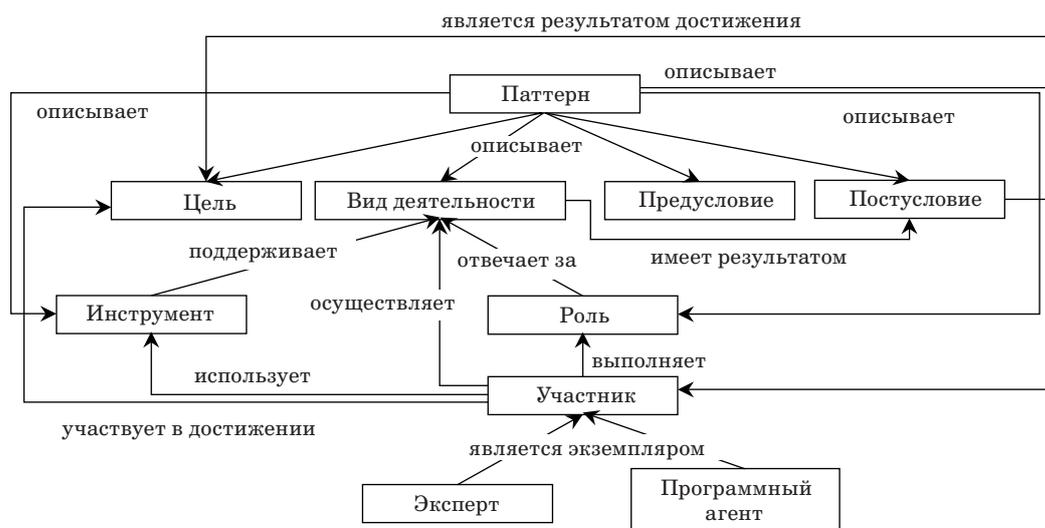
блем и таким образом обеспечивают эффективное совместное поведение участников цифрового сообщества и эффективное принятие решений этим сообществом [10, 11].

### **Концептуальная модель паттерна коллективных действий**

Концептуальная модель паттерна коллективных действий разрабатывалась на основе анализа различных видов паттернов коллективных действий и является обобщением концептуальных моделей таких паттернов. Подробно с результатами анализа можно ознакомиться в работах [12, 13] авторов данной статьи, где вместо паттернов коллективных действий использовался термин паттерны человеко-машинного сотрудничества. Рассматривались паттерны, связанные как с коллективными действиями совместно людей и машин, так и с действиями отдельно групп людей или машин. В результате анализа было выявлено пять видов паттернов: организационный паттерн, паттерн процесса, когнитивный паттерн, паттерн коллаборативной инженерии, паттерн взаимодействия.

Концептуальная модель паттерна коллективных действий [13] (рис. 1) поддерживает следующую семантику. Паттерн предлагает решение, описывающее, каким образом может быть достигнута цель коллективных действий. Множество предусловий определяет возможность использования паттерна. В достижение цели вовлечены участники коллективных действий (участники цифрового сообщества). Такими участниками являются люди и программные агенты. В процессе коллективных действий участники осуществляют деятельность (множество видов деятельности), выполнение которой приводит к достижению цели. Участники действуют в соответствии с ролями, которые они выполняют, и, при необходимости, используют инструменты, поддерживающие определенные виды деятельности. Различные виды деятельности участников приводят к различным результатам (постусловиям видов деятельности). Постусловия видов деятельности, которые ведут к достижению цели коллективных действий, образуют множество постусловий использования паттерна.

Пять видов паттернов предлагают типовые решения для достижения пяти целей: формирования архитектуры среды коллективных действий (организационный паттерн), создания плана коллективных действий для достижения цели (паттерн процесса), создания плана коллективных действий при решении интеллектуальной задачи (когнитивный паттерн), разработки сценария взаимодействия участников (паттерн



■ **Рис. 1.** Концептуальная модель паттерна коллективных действий  
 ■ **Fig. 1.** Conceptual model of collaboration pattern

взаимодействия), установки правил (процедуры) принятия коллективных решений (паттерн коллаборативной инженерии).

Паттерны коллективных действий являются средством управления совместной деятельностью участников цифрового сообщества в процессе решения ими стоящих перед сообществом задач. В отличие от цели, которая появляется на определенных этапах решения задачи и для достижения которой может быть предложено типовое решение, задача — это конкретный вопрос, требующий ответа или решения. В данной статье задачей, стоящей перед сообществом, является задача управления конфигурацией УДС умного города. В статье предполагается, что эта задача решается как задача принятия решений.

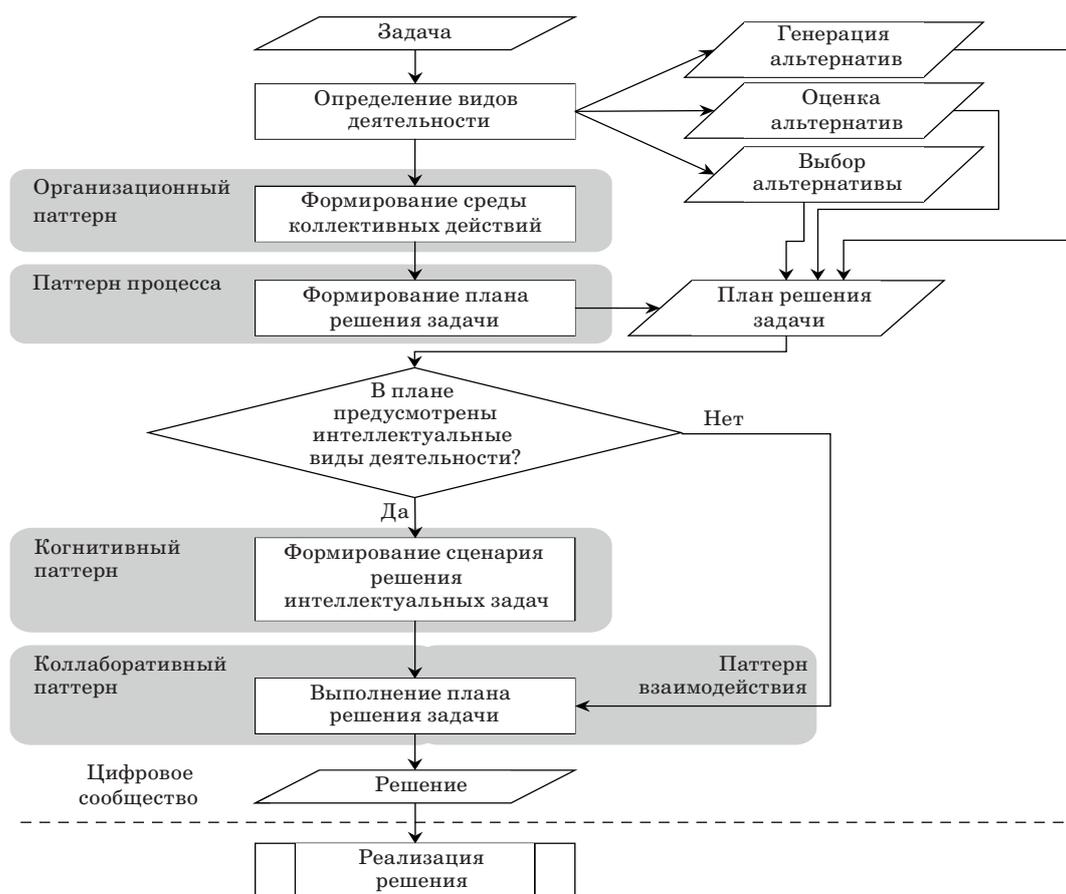
### Обобщенная модель выполнения цифровым сообществом задачи принятия решений на основе использования паттернов коллективных действий

Обобщенная модель выполнения цифровым сообществом стоящей перед ним задачи как задачи принятия решений на основе использования паттернов коллективных действий является контекстно ориентированной. Контекстное представление информации соответствует ее представлению в паттернах коллективных действий. Благодаря этому отсутствует необходимость в дополнительной обработке информации при реализации предлагаемых паттернами решений, а единство терминологии упрощает взаимодействие участников коллективных действий.

В контексте представлена вся информация о текущей ситуации, включая информацию об имеющихся ресурсах (участниках коллективных действий и используемых ими инструментах), выполняемых участниками ролях, осуществляемых ими видах деятельности и задаче. Одной из характеристик задачи является ее статус в данный момент времени. Статус задачи может иметь значения: новая, намечена, запланирована, назначена, в работе, завершена. Совместно или по отдельности статус задачи и выполняемые виды деятельности определяют текущую цель коллективных действий, в соответствии с которой выбирается тот или иной паттерн. Паттерн предлагает типовое решение по достижению текущей цели.

Процесс выполнения задачи на основе паттернов коллективных действий описан в сценарии поддержки принятия решений [13]. В соответствии с этим сценарием задача при появлении имеет статус «Новая». Этот статус означает, что текущей целью является определение видов деятельности, требующихся для решения задачи. Поскольку задача решается как задача принятия решений, такими видами деятельности являются идентификация цели, генерация альтернатив, оценка альтернатив и выбор альтернативы (принятие решения) [14, 15].

После того как виды деятельности определены, задача приобретает статус «Намечена». Дальнейшая деятельность участников цифрового сообщества, нацеленная на выполнение задачи, поддерживается паттернами коллективных действий (рис. 2). Условия выбора этих паттернов приведены в табл. 1.



■ **Рис. 2.** Обобщенная модель выполнения задачи принятия решений на основе паттернов коллективных действий  
 ■ **Fig. 2.** Generalized model of fulfilling decision-making problem based on collaboration patterns

■ **Таблица 1.** Условия выбора паттернов коллективных действий  
 ■ **Table 1.** Conditions for choice of collaboration patterns

Предусловия	Паттерн	Цель	Проблема	Постусловия
Статус задачи: «Намечена»	Организационный	Формирование архитектуры среды коллективных действий	Определение множества компонентов среды коллективных действий	Архитектура среды коллективных действий Статус задачи: «Запланирована»
Статус задачи: «Запланирована»	Процесса	Формирование плана решения задачи	Планирование деятельности участников	План коллективных действий Статус задачи: «Назначена»
Статус задачи: «В работе» Характер вида деятельности: интеллектуальный	Когнитивный	Формирование плана решения интеллектуальной задачи	Организация процесса решения интеллектуальной задачи	Сценарий решения интеллектуальной задачи
Статус задачи: «В работе» Запрос на принятие решений	Коллаборативной инженерии	Установка правила (процедуры) коллективного принятия решений	Выбор и согласование правила коллективного принятия решений	Правило коллективного принятия решений
Статус задачи: «В работе» Запрос на взаимодействие	Взаимодействия	Формирование сценария взаимодействия участников	Определение инструментов и коммуникативных актов для взаимодействия	Сценарий взаимодействия участников

## Сценарий решения цифровой сообществом задачи предоставления рекомендаций по конфигурации УДС умного города

Задача управления конфигурацией УДС умного города решается в рамках мероприятий по уменьшению количества дорожно-транспортных происшествий (ДТП). Эта задача ставится программным агентом системы «умный город», который на основании анализа данных о ДТП на определенном участке дороги формирует перечень мест концентрации ДТП за отчетный год. Местом концентрации ДТП называется участок дороги, улицы, не превышающий 1000 метров вне населенного пункта или 200 метров в населенном пункте, либо пересечение дорог, улиц, где в течение отчетного года произошло три и более ДТП одного вида или пять и более ДТП независимо от их вида, в результате которых погибли или были ранены люди (Федеральный закон от 10.12.1995 № 196-ФЗ (ред. от 08.08.2024) «О безопасности дорожного движения» (с изм. и доп., вступ. в силу с 01.03.2025): 196-ФЗ. РФ, 1995, 45 с.). Например, за 2024 г. в Санкт-Петербурге выявлено 91 место концентрации ДТП, среди которых 10 выделены как места с самой высокой концентрацией однотипных ДТП (Комплексная схема организации дорожного движения Санкт-Петербурга. Санкт-Петербург, Правительство Санкт-Петербурга. Комитет по развитию транспортной инфраструктуры Санкт-Петербурга, 2024. 201 с.). Для принятия решения о проведении мероприятий в части реконфигурации технических средств организации дорожного движения, направленных на снижение количества ДТП, информация о данных местах передается в человеко-машинную среду коллективного интеллекта (ЧМСКИ) [16]. Данная среда оказывает технологическую поддержку, обеспечивая возможность взаимодействия участников, поддерживая их интероперабельность, а также предоставляя механизмы самоорганизации и средства для выполнения сложных вычислительных операций, которые реализуются агентами.

Человеко-машинная среда коллективного интеллекта формирует для участников цифрового сообщества задачу предоставления рекомендаций по конфигурации УДС умного города в конкретном месте концентрации ДТП. Эта задача решается в рамках регулярного обследования, направленного на выявление и устранение мест концентрации ДТП и чаще всего предусматривает реконфигурацию технических средств организации дорожного движения.

При поступлении задачи ей присваивается статус «Новая», и в соответствии с обобщенной

моделью выполнения задачи принятия решений (см. рис. 2) определяются виды деятельности, требующиеся для ее решения как задачи поддержки принятия решений. Такими видами являются генерация множества возможных рекомендаций по реконфигурации технических средств организации дорожного движения, направленных на снижение количества ДТП в заданном месте, оценка этих рекомендаций относительно критериев, принятых цифровым сообществом, и принятие решения о предоставляемой рекомендации по снижению количества ДТП. На интервале от момента поступления задачи до момента определения видов деятельности в контексте представлено значение статуса задачи – «Новая» и текущая цель – «Определение видов деятельности». После того как эти виды определены, в контексте появляются перечисленные выше виды деятельности.

По завершении определения видов деятельности задаче присваивается статус «Намечена», который передается в контекст. Благодаря единому представлению информации в контексте и паттернах наличие в контексте данного статуса сразу позволяет обратиться к организационному паттерну, т. е. инициировать деятельность по созданию архитектуры среды коллективных действий. Соответственно, в контексте появляется цель «Формирование архитектуры среды коллективных действий». В результате применения указанного паттерна определяются ресурсы среды, выполняемые средой функции и роли участников коллективных действий.

Ресурсами, включенными в архитектуру среды коллективных действий (рис. 3), являются участники цифрового сообщества (агенты и эксперты), а также ЧМСКИ, которая рассматривается как инструмент, обеспечивающий технологическую поддержку участников. Взаимодействие агентов с данной средой осуществляется через протокол REST API, а экспертов – через пользовательский веб-интерфейс.

Группа агентов включает в себя *аналитика происшествий* (предоставляет информацию о ДТП, произошедших в заданном месте, и анализирует карточки учета ДТП (Приложение А к Распоряжению Росавтодора № 853-р от 12 мая 2015 г. (ред. от 31.01.2017) «Об издании и применении ОДМ 218.6.015-2015 «Рекомендации по учету и анализу дорожно-транспортных происшествий на автомобильных дорогах Российской Федерации», 2017. 106 с.); множество *аналитиков контекста* (получают данные из погодного архива и системы контроля дорожных фондов (СКДФ)); *менеджера контекста* (осуществляет расширение контекста информацией из СКДФ о дороге, на которой расположено место концентрации ДТП, знаках, установленных в месте

концентрации ДТП, и погоде в момент каждого ДТП); *рекомендательного агента* (предлагает рекомендации мероприятий по снижению и профилактике аварийности); *предписывающего агента* (предлагает критерии оценки альтернатив, типовые регламенты решения задач, документы для ознакомления).

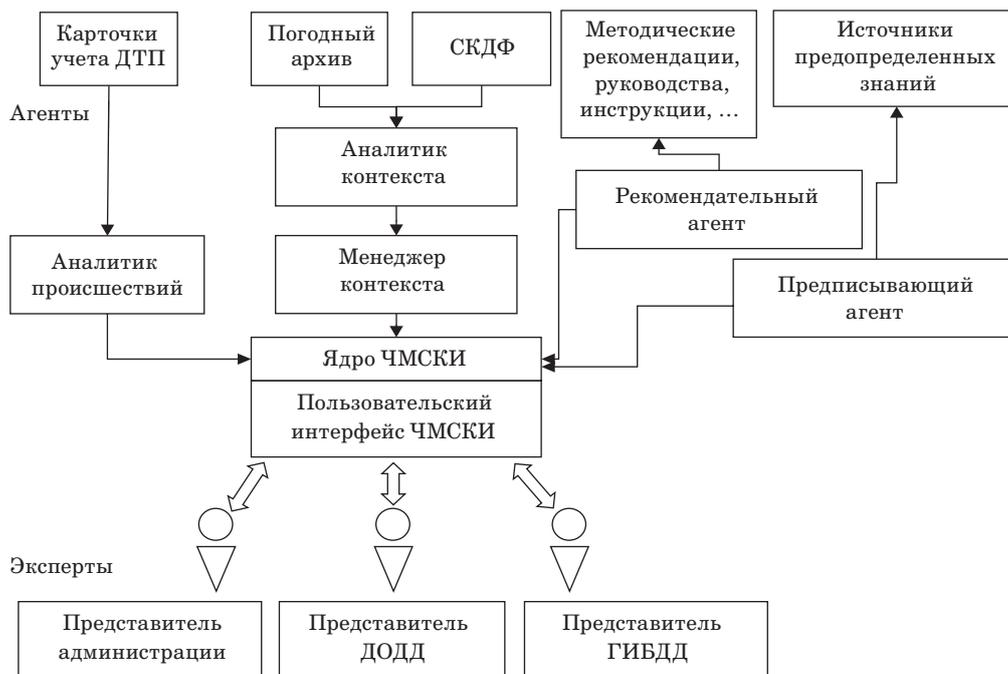
Группа экспертов включает в себя представителей исполнительных органов государственной власти различных уровней, ведомств, предприятий, учреждений, организаций, обеспечивающих безопасность дорожного движения. В рамках предлагаемого сценария в качестве экспертов рассматриваются представители рабочей группы Комиссии по обеспечению безопасности дорожного движения при губернаторе Санкт-Петербурга, в которую входят специалисты Управления ГИБДД ГУ МВД России по г. Санкт-Петербургу и Ленинградской области (*представитель ГИБДД*), Комитета по развитию транспортной инфраструктуры (*представитель администрации*), СПб ГКУ «Дирекция по организации дорожного движения Санкт-Петербурга» (*представитель ДОДД*).

Для участников коллективных действий определены следующие роли: *ЛПР* (лицо, принимающее решения), *аналитик* (занимается аналитическими исследованиями, систематизацией и обобщением информации и данных), *специалист по организации рабочего процесса* (занимается разработкой сценариев решения интеллектуальных задач), *помощник* (предоставляет

сервисы для эксперта) и *посредник* (осуществляет передачу информации в контекст). Участники коллективных действий могут выполнять одну или несколько из приведенных здесь ролей. Сами роли и распределение их между участниками на рис. 3 не показаны, так как назначение участника на роль осуществляется при планировании процесса решения задачи. Далее в тексте имена участников пишутся с прописной буквы, а названия выполняемых ими ролей – со строчной.

В множество функций, исполняемых средой коллективных действий, входят функции по выполнению видов деятельности, которые были определены для поступившей задачи (генерация множества возможных рекомендаций, оценка рекомендаций, принятие решения о предоставляемой рекомендации), взаимодействию участников и предоставлению информации.

Формирование архитектуры среды коллективных действий завершается тем, что ее компоненты отображаются в контексте, а задаче присваивается статус «Запланирована», который также передается в контекст. В соответствии с данным статусом участники переходят к разработке плана решения задачи, в контексте появляется цель «Формирование плана решения задачи». Поскольку формирование рекомендаций происходит в рамках регулярного (планового) обследования, для планирования деятельности участников *предписывающий агент* предлагает типовой план, ранее построенный при помощи



■ **Рис. 3.** Ресурсы среды коллективных действий  
 ■ **Fig. 3.** Resources of collaboration environment

паттерна процесса и используемый для решения подобных задач в ЧМСКИ. Конкретизация типового плана применительно к сформированной архитектуре среды коллективных действий показана в табл. 2. Этот план передается в контекст. В таблице столбец с номером обозначает порядок вида деятельности в запланированной последовательности действий. Позиции «Взаимодействие» и «Изменение контекста» номеров не имеют, поскольку они могут потребоваться при выполнении любого вида деятельности.

После создания плана задаче присваивается статус «Назначена», участники коллективных действий приступают к его реализации, и статус задачи меняется на «В работе». Статусы задач и виды деятельности, которые участники выполняют по мере реализации плана, и их результаты, отображаются в контексте. Далее во избежание ненужного увеличения объема статьи в описании не указывается, что конкретный вид деятельности и полученный результат отображены в контексте, поскольку это и так понятно.

На этапе сбора информации о задаче (№ 1 в плане решения задачи (см. табл. 2)) *аналитик происшествий* предоставляет участникам координаты места ДТП, например пересечение набережной Обводного канала и Московского проспекта (рис. 4), и статистическую сводку за отчетный период (2023–2024 гг.) для исследования. Согласно предоставленной сводке, за этот период на данном участке произошло шесть ДТП с общим числом участников – 8. *Аналитик контекста* предоставляет дополнительную информацию о погоде во всех местах ДТП, типе и состоянии дорожного покрытия,



Тип ДТП:

- × столкновение
- наезд на пешехода
- наезд на велосипедиста

■ **Рис. 4.** Карта с визуализацией места концентрации ДТП

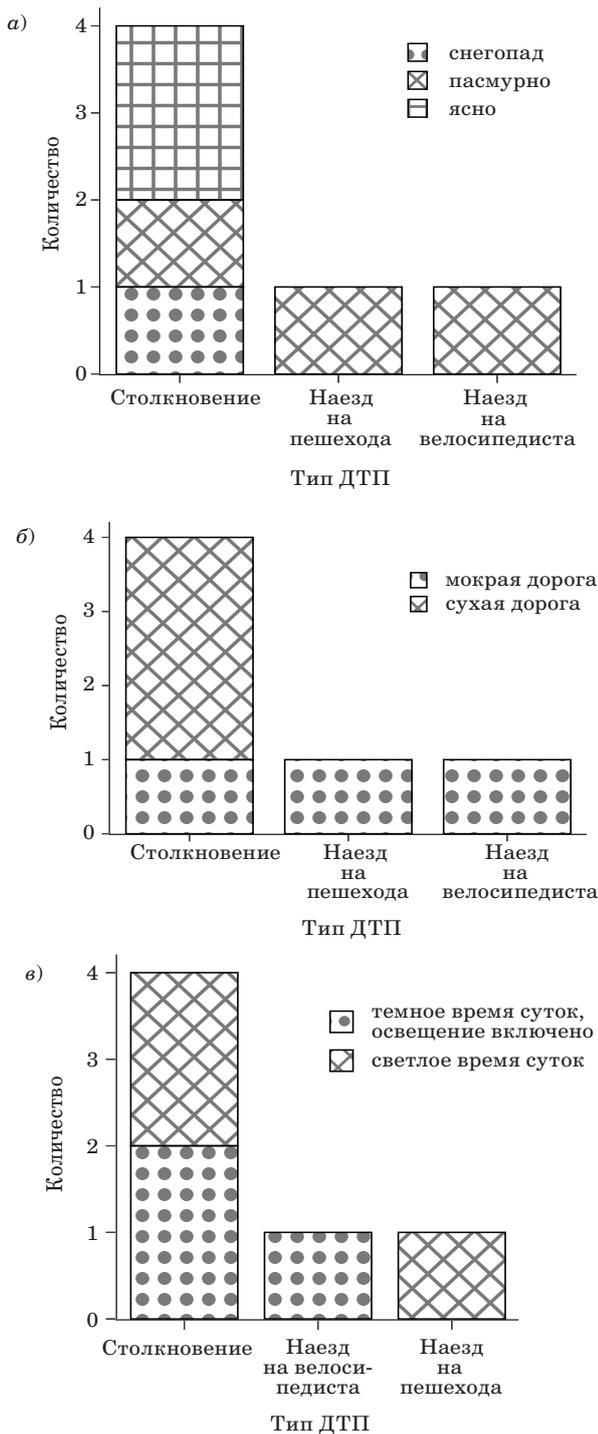
■ **Fig. 4.** Map visualizing a place of traffic accidents concentration

пропускной способности дороги. На основе этой информации проводится статистический анализ, результаты которого представлены гистограммами (рис. 5, а–в), свидетельствующими о том, что основным видом ДТП является столкновение, преимущественно в пасмурную погоду, в снегопад и в темное время суток, при этом искусственное освещение включено. Основными нарушителями являются водители

■ **Таблица 2.** План решения задачи управления конфигурацией УДС

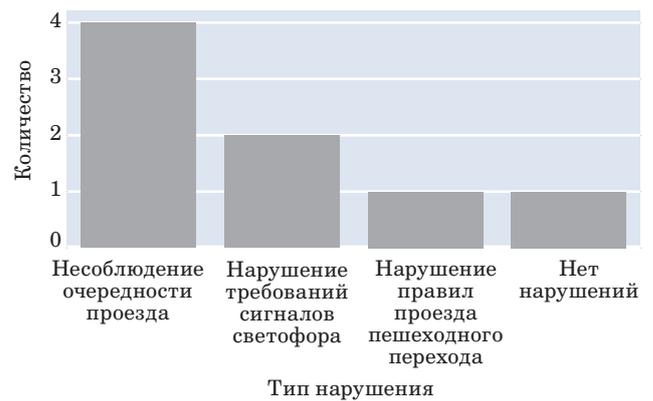
■ **Table 2.** Plan for solving problem of configuration management of street and road network

№	Вид деятельности	Роль	Участник	Инструмент
1	Сбор информации о задаче	Аналитик	Аналитик происшествий Аналитики контекста	ЧМСКИ, REST API
2	Генерация множества рекомендаций по снижению ДТП	Аналитик	Рекомендательный агент	ЧМСКИ, REST API
3	Оценка рекомендаций (интеллектуальная задача)	Специалист по организации рабочего процесса	Представитель администрации	ЧМСКИ, веб-интерфейс
		Помощник	Предписывающий агент	
		ЛПР	Эксперты	
4	Принятие решения	ЛПР	Эксперты	ЧМСКИ, веб-интерфейс
–	Взаимодействие	Инициатор Исполнитель	Все участники	ЧМСКИ, веб-интерфейс, REST API
–	Изменение контекста	Посредник	Менеджер контекста	ЧМСКИ, REST API



■ **Рис. 5.** Результаты статистической обработки информации о месте концентрации ДТП по типу погоды (а); по состоянию дороги (б); по типу освещения (в)  
 ■ **Fig. 5.** Results of statistics processing information on place of traffic accidents concentration by weather type (a); by road conditions (b); by lighting type (в)

транспортных средств, а основными типами нарушений являются «несоблюдение очередности проезда» и «нарушение требований сигналов светофора» (рис. 6). Согласно данным СКДФ,



■ **Рис. 6.** Частота нарушений правил дорожного движения в месте концентрации ДТП  
 ■ **Fig. 6.** Frequency of traffic violations in place of traffic accidents concentration

на этом участке пересекаются две дороги, одна из которых – набережная Обводного канала с пропускной способностью 302 400 автомобилей в сутки, среднесуточной интенсивностью движения 28 750 автомобилей в сутки, ограничением скорости 60 км/ч, асфальтобетонным покрытием, а другая – Московский проспект с пропускной способностью 368 871 автомобиль в сутки, среднесуточной интенсивностью движения 61 447 автомобилей в сутки, ограничением скорости 60 км/ч, асфальтобетонным покрытием. Состояние дорог соответствует нормативному.

После того как информация о задаче собрана, *рекомендательный агент* приступает к деятельности по генерации рекомендаций по снижению ДТП. Он пользуется «Методическими рекомендациями по выбору эффективных некапиталоемких мероприятий по снижению аварийности в местах концентрации ДТП на автомобильных дорогах общего пользования» (ОДМ 218.6.025–2017, Москва, 2020. 46 с.) и «Руководством по устранению и профилактике возникновения участков концентрации ДТП при эксплуатации автомобильных дорог» (ОДМ 218.4.004-2009, Москва, 2009. 94 с.). На основе этих документов агент предлагает несколько рекомендаций:

- а) пересмотр установленных дорожных знаков, установка дополнительных знаков, увеличивающих осведомленность водителей о схеме проезда перекрестка;
- б) изменение режима работы светофоров для разделения потоков пешеходов и встречных потоков автомобилей;
- в) установка камер для фото- и видеофиксации и автоматическая обработка нарушений.

Предлагаемые рекомендации подлежат оценке. В плане решения задачи управления конфигурацией УДС оценка рекомендаций определена как интеллектуальная задача, поэтому оценку

выполняют эксперты. Процесс решения интеллектуальной задачи начинается с создания плана ее решения. Такой план разрабатывается *представителем администрации* при поддержке *предписывающего агента*. Этот агент на основе когнитивного паттерна [17] предлагает рекомендуемую последовательность действий по решению задачи, а *представитель администрации* уточняет, изменяет и конкретизирует данную последовательность.

*Предписывающий агент* предложил приведенную далее очередность действий: 1) генерация (вид деятельности, результатом которой является множество предложений  $G$ , с которыми согласны все эксперты); 2) уточнение (вид деятельности, результат которой свидетельствует о том, что все эксперты адекватно понимают выдвинутые ими предложения); 3) сокращение (вид деятельности, результатом которой является множество предложений  $R$ , отобранных из множества  $G$  и представляющих, по мнению экспертов, интерес); 4) уточнение; 5) достижение консенсуса (вид деятельности, в процессе которой отслеживается количество экспертов, поддерживающих одно и то же предложение из множества  $R$ ).

*Представитель администрации* отказывается от деятельности по уточнению, так как считает, что эксперты образуют малую группу, хорошо понимают друг друга, и необходимость в дополнительной деятельности по уточнению высказываний друг друга отсутствует. В результате он конкретизирует последовательность действий по решению задачи оценки рекомендаций следующим образом: 1) генерация – предоставление множества критериев для оценки; 2) достижение консенсуса – выбор критерия; 3) сокращение – выбор рекомендации относительно выбранного критерия. План рабочего процесса решения задачи оценки рекомендаций приведен в табл. 3. Роль *помощника* в этом плане выполняет *предписывающий агент*, который предоставляет экспертам множество критериев оценки.

В процессе реализации плана решения задачи по оценке рекомендаций эксперты выбирают критерий целесообразности. Деятельность,

связанная с выбором рекомендации (сокращением количества рекомендаций), соответствует деятельности по принятию решения. Для ее осуществления используется коллаборативный паттерн [18], который позволяет установить правило принятия решений на основе анализа контекста принятия решений и плюсов и минусов различных правил принятия решений.

Перед использованием коллаборативного паттерна определяется ответственный за выбор правила принятия решений, а участникам принятия решений разъясняется процедура выбора такого правила и объясняется процесс принятия решений в соответствии с правилами.

В рассматриваемом здесь примере ответственным за выбор правила принятия решений считается *представитель администрации*, поскольку именно администрация отвечает за продвижение рекомендаций, которые будут предложены цифровым сообществом. Перед началом процедуры выбора правила *предписывающий агент* предлагает *представителю администрации* и остальным экспертам для ознакомления документ, описывающий процедуру выбора правила принятия решений и процесс принятия решений в соответствии с правилами. В соответствии с паттерном ответственный за выбор правила принятия решений выбирает из множества правил правило, которое он считает подходящим, а затем согласовывает свой выбор с остальными участниками. После того как правило согласовано, оно становится правилом, которым участники пользуются при принятии решений.

*Представитель администрации* выбирает правило консенсуса. Контекст применения данного правила в паттерне описан как «состояние взаимного согласия между участниками, при котором интересы отдельных участников направлены на удовлетворение интересов коллектива». В качестве плюсов данного правила паттерн указывает: правило укрепляет доверие; обеспечивает высокий уровень поддержки принятого решения и приверженности к нему; рассматривает последствия принятого решения; способствует обучению, так как требует внимательного слушания и изучения; повышает рациональность

■ **Таблица 3.** План рабочего процесса решения интеллектуальной задачи по оценке рекомендаций

■ **Table 3.** Plan of workflow for solving intellectual task on evaluation of recommendations

№	Вид деятельности	Цель	Роль	Участник	Инструмент
1	Генерация	Предоставление множества критериев оценки	Помощник	Предписывающий агент	ЧМСКИ, REST API
2	Достижение консенсуса	Выбор критерия, с которым согласна группа экспертов	ЛПР	Эксперты	ЧМСКИ, веб-интерфейс
3	Сокращение	Выбор рекомендации	ЛПР	Эксперты	ЧМСКИ, веб-интерфейс

принимаемого решения. Минусы правила: занимает много времени; участники принятия решений должны обладать опытом и знаниями в рассматриваемой области; качество решения может снизиться, если участники не располагают всей необходимой информацией.

Выбранное *представителем администрации* правило, его описание, плюсы и минусы были переданы остальным экспертам на предмет получения от них мнения, согласны они пользоваться правилом консенсуса при принятии решений или нет. Все эксперты выразили свое согласие. После чего они используют это правило при выборе рекомендации относительно критерия целесообразности из множества рекомендаций «а, б, в», которые были предложены *рекомендательным агентом*.

Относительно критерия целесообразности *представители ГИБДД и ДОДД* считают, что рекомендация «а» о пересмотре установленных дорожных знаков, установке дополнительных знаков, увеличивающих осведомленность водителей о схеме проезда перекрестка, нецелесообразна, так как в случае установки дополнительных знаков схема проезда усложнится относительно текущей, что увеличит количество столкновений при проезде перекрестка. Таким образом, данная рекомендация отклоняется.

Касательно рекомендаций «б» и «в» — изменение режима работы светофора и установка дополнительных устройств фото- и видеofиксации нарушений — *представители администрации и ГИБДД* решают, что имеет смысл рассматривать обе рекомендации (рис. 7). *Представитель администрации* предлагает внедрение в рассматриваемом месте системы «умный перекресток»,



■ **Рис. 7.** Предлагаемые места установки камер фото- и видеofиксации на рассматриваемом месте концентрации ДТП

■ **Fig. 7.** Proposed locations for installation of photo and video cameras at the considered place of traffic accidents concentration

которая обнаруживает транспорт с помощью заложенных в дорожное полотно индуктивных петлевых детекторов и управляет светофорами в адаптивном режиме. По мнению данного эксперта, установка такой системы в рассматриваемом месте наиболее целесообразна.

Остальные эксперты согласны с оценкой *представителя администрации*, и, таким образом, принимается рекомендация о внедрении системы «умный перекресток» с автоматической регулировкой режима работы светофора и фото- и видеofиксацией нарушений.

При любом взаимодействии участников цифрового сообщества применяется паттерн взаимодействия. В таких случаях ЧМСКИ конкретизирует соответствующие элементы паттерна (последовательность элементарных взаимодействий участников (запрос/ответ)); контент, передаваемый при взаимодействии; роли участников взаимодействий и используемые средства коммуникации (инструменты), а также самих участников взаимодействия в связи с выполняемыми ими ролями. В рассматриваемом паттерне для участников определены две роли: *инициатор* взаимодействия и *исполнитель* (адресат взаимодействия). Инструментом взаимодействия является платформа ЧМСКИ.

## Заключение

Рассмотрена проблема управления конфигурацией социотехнической системы на примере решения задачи управления конфигурацией улично-дорожной сети умного города. В результате исследований по проблеме предложены концептуальная модель паттернов коллективных действий, обобщенная модель выполнения задачи принятия решений на основе паттернов коллективных действий и сценарий решения цифровым сообществом задачи предоставления рекомендаций по изменению конфигурации УДС умного города.

Концептуальная модель паттернов коллективных действий является обобщением пяти видов паттернов, где каждый вид паттерна предлагает типовое решение для конкретной задачи, регулярно возникающей в процессе коллективных действий. Данная модель унифицирует концепты, которые присутствуют в спецификациях всех видов паттернов, благодаря чему упрощается обработка контекстной информации и взаимодействие участников цифрового сообщества.

Обобщенная модель выполнения задачи принятия решений на основе паттернов коллективных действий является моделью совместного решения любой прикладной задачи как задачи принятия решений. Она поддерживает кон-

текстный выбор паттернов в процессе коллективного решения задачи на основе соответствий между контекстной информацией и значениями предусловий использования паттернов, а также зависимостей между пред- и постусловиями применения паттернов.

Сценарий решения задачи предоставления рекомендаций по управлению конфигурацией УДС умного города реализован на примере решения задачи цифровым сообществом, которое состоит из рабочей группы Комиссии по обеспечению безопасности дорожного движения при администрации города и программных агентов и деятельность участников которого поддерживается платформой человеко-машинной среды

коллективного интеллекта. Сценарий демонстрирует применимость разработанных моделей к реальным условиям.

### Финансовая поддержка

Исследование выполнено в рамках бюджетной темы FFZF-2025-0003 «Разработка теоретических и технологических основ анализа неструктурированных данных и многомодального взаимодействия пользователей, интеллектуальной поддержки принятия решений в цифровых сообществах».

### Литература

1. Видясова Л. А. Концептуализация понятия «умный город»: социотехнический подход. *International Journal of Open Information Technologies*, 2017, т. 5, № 11, с. 52–57. EDN: XEOHNC. <http://injoit.org/index.php/j1/article/view/506/480> (дата обращения: 30.01.2025).
2. Сергеев С. Социотехнические системы с искусственным интеллектом: вопросы теории и методологии. *Институт психологии Российской академии наук. Организационная психология и психология труда*, 2022, вып. 7, № 1, с. 4–23. doi:10.38098/ipran.opwp\_2022\_22\_1\_001. <http://work-orgpsychology.ru/engine/documents/document750.pdf> (дата обращения: 30.06.2025).
3. Yaqoob I., Salah K., Jayarama R., Omar M. Metaverse applications in smart cities: Enabling technologies, opportunities, challenges, and future directions. *Internet of Things*, 2023, vol. 23, Article 100884. doi:10.1016/j.iot.2023.100884
4. Pandya S., Srivastava G., Jhaveri R., Babu M. R., Bhattacharya S., Maddikunta P. K. R., Mastorakis S., Piran Md. J., Gadekallu T. R. Federated learning for smart cities: A comprehensive survey. *Sustainable Energy Technologies and Assessments*, 2023, vol. 55, Article 102987. doi:10.1016/j.seta.2022.102987
5. *Digital Communities. Sustainability Directory*, 2024. <https://sustainability-directory.com/term/definition/digital-communities/> (дата обращения: 11.01.2025).
6. Xu J. Digital community management mobile information system based on edge computing. *Mobile Information Systems*, 2021, vol. 2021, iss. 8, pp. 1–11. doi:10.1155/2021/5842291
7. Tupayachi J., Xu H., Omitaomu O. A., Camur M. C., Sharmin A., Li X. Towards next-generation urban decision support systems through AI-powered construction of scientific ontology using large language models – a case in optimizing intermodal freight transportation. *Smart Cities*, 2024, vol. 7, no. 5, pp. 2392–2421. doi:10.3390/smartcities7050094
8. Kutty A. A., Kucukvar M., Onat N. C., Ayvaz B., Abdella G. M. Measuring sustainability, resilience and livability performance of European smart cities: A novel fuzzy expert-based multi-criteria decision support model. *Cities*, 2023, vol. 137, Article 104293. doi:10.1016/j.cities.2023.104293
9. Olaniyi O. O., Okunleye O. J., Olabanji S. O. Advancing data-driven decision-making in smart cities through big data analytics: A comprehensive review of existing literature. *Current Journal of Applied Science and Technology*, 2023, vol. 42, no. 25, pp. 10–18. doi:10.9734/cjast/2023/v42i254181
10. Fan S., Li X., Zhao J. L. Collaboration process pattern approach to improving teamwork performance: A data mining-based methodology. *INFORMS Journal on Computing*, 2017, vol. 29, no. 3, pp. 438–456. doi:10.1287/ijoc.2016.0739
11. Richards D., Cowell-Butler J. Decisions within human-machine teaming: The introduction of decision strings. *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, IEEE, 2022, pp. 1–7. doi:10.1109/ICHMS56717.2022.9980668
12. Смирнов А. В., Левашова Т. В. Паттерны человеко-машинного сотрудничества в системах поддержки принятия решений. *Искусственный интеллект и принятие решений*, 2024, № 2, с. 3–17. doi:10.14357/20718594240201
13. Смирнов А. В., Левашова Т. В. Онтология паттернов человеко-машинного сотрудничества для поддержки принятия решений. *Онтология проектирования*, 2024, т. 14, № 3, с. 421–439. doi:10.18287/2223-9537-2024-14-3-421-439
14. Simon H. *The New Science of Decision-Making*. New York, Harper and Row, 1960. 50 p.
15. Simon H. Rational decision making in business organizations. *American Economic Association*, 1979, vol. 69, no. 4, pp. 493–513.
16. Смирнов А. В., Пономарев А. В., Шилов Н. Г., Левашова Т. В., Тесля Н. Н. Концепция построения коллаборативных систем поддержки принятия решений: подход и архитектура платформы. *Инфор-*

матика и автоматизация, 2024, вып. 23, № 4, с. 1139–1172. doi:10.15622/ia.23.4.8

17. Vreede G. J. De, Kolfshoten G. L., Briggs R. O. ThinkLets: a collaboration engineering pattern language. *International Journal of Computer Applications in Technology*, 2006, vol. 25, no. 2/3, pp. 140–154. doi:10.1504/IJCAT.2006.009064

18. Gottesdiener E. Decide how to decide. *Software Development Magazine*, 2001, vol. 9, no. 1. <https://www.ebgconsulting.com/Pubs/Articles/DecideHowToDecide-Gottesdiener.pdf> (дата обращения: 11.01.2025).

UDC 004.89:004.048+004.5

doi:10.31799/1684-8853-2025-4-13-25

EDN: ZOZUHI

### Management of smart city road network configuration: a scenario based on collaboration patterns of decision-making participants

A. V. Smirnov<sup>a</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0001-8364-073X](https://orcid.org/0000-0001-8364-073X)

T. V. Levashova<sup>a</sup>, PhD, Tech., Senior Researcher, [orcid.org/0000-0002-1962-7044](https://orcid.org/0000-0002-1962-7044), [tatiana.levashova@iias.spb.su](mailto:tatiana.levashova@iias.spb.su)

N. N. Teslya<sup>a</sup>, PhD, Tech., Senior Researcher, [orcid.org/0000-0003-0619-8620](https://orcid.org/0000-0003-0619-8620)

<sup>a</sup>St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

**Introduction:** In the process of joint activities of participants of digital communities consisting of people and software agents, repetitive problems often arise. To organize purposeful joint activities of the participants collaboration patterns that provide reusable solutions for recurring problems can be used. **Purpose:** To develop models for the effective organization of purposeful activities of participants of digital communities based on the collaborative patterns in the process of jointly solving the problem of management of a sociotechnical system and to propose a scenario for making recommendations on the management of a configuration of the smart city road network as a kind of sociotechnical system. **Results:** We develop a conceptual model of a collaboration pattern which facilitates contextual information processing and interactions of the participants of a digital community due to the homogeneous representation of information used in the specifications of patterns of different types. We further develop a generalized decision-making model based on collaboration patterns which supports the choice of patterns in the process of joint activities within digital participation. We propose a scenario for making recommendations on managing the configuration of a smart city street and road network by the participants of a digital community which confirms the adequacy of the developed models. **Practical relevance:** The research results contribute to the problem of the management of sociotechnical system configurations. They provide models for the effective organization of purposeful joint activities of participants of digital communities in the process of their collaborative solving the configuration management problem as a decision-making problem. These results can be used, for instance, to make recommendations on the configuration management for such systems as a smart city, an airport and others.

**Keywords** – smart city, sociotechnical system, digital community, street and road network, decision support, collaboration patterns.

**For citation:** Smirnov A. V., Levashova T. V., Teslya N. N. Management of smart city road network configuration: a scenario based on collaboration patterns of decision-making participants. *Informatsionno-upravliaushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 13–25 (In Russian). doi:10.31799/1684-8853-2025-4-13-25, EDN: ZOZUHI

#### Financial support

The research is due to the State Research Project FFZF-2025-0003 “The development of theoretical and technological foundations for the analysis of unstructured data and multimodal user interactions, and for intellectual decision support in digital communities”.

#### References

- Vidiasova L. Conceptualization of the “Smart City” concept: Socio-technical approach. *International Journal of Open Information Technologies*, 2017, vol. 5, no. 11, pp. 52–57. Available at: <http://injoit.org/index.php/j1/article/view/506/480> (accessed 30 January 2025) (In Russian).
- Sergeev S. Sociotechnical systems with artificial intelligence: Questions of theory and methodology. *Institute of Psychology of the Russian Academy of Sciences. Organizational Psychology and Labor Psychology*, 2022, vol. 7, no. 1, pp. 4–23. doi:10.38098/ipran.opwp\_2022\_22\_1\_001. Available at: <http://work-org-psychology.ru/engine/documents/document750.pdf> (accessed 30 January 2025) (In Russian).
- Yaqoob I., Salah K., Jayarama R., Omar M. Metaverse applications in smart cities: Enabling technologies, opportunities, challenges, and future directions. *Internet of Things*, 2023, vol. 23, Article 100884. doi:10.1016/j.iot.2023.100884
- Pandya S., Srivastava G., Jhaveri R., Babu M. R., Bhat-tacharya S., Maddikunta P.K.R., Mastorakis S., Piran Md. J., Gadekallu T. R. Federated learning for smart cities: a comprehensive survey. *Sustainable Energy Technologies and Assessments*, 2023, vol. 55, Article 102987. doi:10.1016/j.seta.2022.102987
- Digital Communities. Sustainability Directory*, 2024. Available at: <https://sustainability-directory.com/term/definition/digital-communities/> (accessed 11 January 2025).
- Xu J. Digital community management mobile information system based on edge computing. *Mobile Information Systems*, 2021, vol. 2021, iss. 8, pp. 1–11. doi:10.1155/2021/5842291
- Tupayachi J., Xu H., Omitaomu O. A., Camur M. C., Sharmin A., Li X. Towards next-generation urban decision support systems through AI-powered construction of scientific ontology using large language models – a case in optimizing intermodal freight transportation. *Smart Cities*, 2024, vol. 7, no. 5, pp. 2392–2421. doi:10.3390/smartsities7050094
- Kutty A. A., Kucukvar M., Onat N. C., Ayvaz B., Abdella G. M. Measuring sustainability, resilience and livability performance of European smart cities: A novel fuzzy expert-based multi-criteria decision support model. *Cities*, 2023, vol. 137, Article 104293. doi:10.1016/j.cities.2023.104293
- Olaniyi O. O., Okunleye O. J., Olabanji S. O. Advancing data-driven decision-making in smart cities through big data analytics: A comprehensive review of existing literature. *Current Journal of Applied Science and Technology*, 2023, vol. 42, no. 25, pp. 10–18. doi:10.9734/cjast/2023/v42i254181

10. Fan S., Li X., Zhao J. L. Collaboration process pattern approach to improving teamwork performance: A data mining-based methodology. *INFORMS Journal on Computing*, 2017, vol. 29, no. 3, pp. 438–456. doi:10.1287/ijoc.2016.0739
11. Richards D., Cowell-Butler J. Decisions within human-machine teaming: The introduction of decision strings. *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, IEEE, 2022, pp. 1–7. doi:10.1109/ICHMS56717.2022.9980668
12. Smirnov A., Levashova T. Patterns of human-machine collaboration in decision support systems. *Scientific and Technical Information Processing*, 2024, no. 2, pp. 3–17 (In Russian). doi:10.14357/20718594240201
13. Smirnov A., Levashova T. Ontology of human-machine collaboration patterns for decision support. *Ontology of Designing*, 2024, vol. 14, no. 3, pp. 421–439 (In Russian). doi:10.18287/2223-9537-2024-14-3-421-439
14. Simon H. *The New Science of Decision-Making*. New York, Harper and Row, 1960. 50 p.
15. Simon H. Rational decision making in business organizations. *American Economic Association*, 1979, vol. 69, no. 4, pp. 493–513.
16. Smirnov A., Ponomarev A., Shilov N., Levashova T., Teslya N. A conception of collaborative decision support systems: Approach and platform architecture. *Informatics and Automation*, 2024, vol. 23, no. 4, pp. 1139–1172 (In Russian). doi:10.15622/ia.23.4.8
17. Vreede G. J. De, Kolfshoten G. L., Briggs R. O. ThinkLets: A collaboration engineering pattern language. *International Journal of Computer Applications in Technology*, 2006, vol. 25, no. 2/3, pp. 140–154. doi:10.1504/IJCAT.2006.009064
18. Gottesdiener E. Decide how to decide. *Software Development Magazine*, 2001, vol. 9, no. 1. Available at: <https://www.ebg-consulting.com/Pubs/Articles/DecideHowToDecide-Gottesdiener.pdf> (accessed 11 January 2025).

### УВАЖАЕМЫЕ АВТОРЫ!

Научная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы регистрируетесь на сайте НЭБ (<http://elibrary.ru/defaultx.asp>), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющих в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.



## Гибридный метод синхронизации времени в распределенных системах

Т. М. Татарникова<sup>а</sup>, доктор техн. наук, профессор, [orcid.org/0000-0002-6419-0072](https://orcid.org/0000-0002-6419-0072), [tm-tatarn@yandex.ru](mailto:tm-tatarn@yandex.ru)

Е. Д. Архипцев<sup>б</sup>, аспирант, [orcid.org/0009-0006-3001-953X](https://orcid.org/0009-0006-3001-953X)

<sup>а</sup>Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

<sup>б</sup>Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», Профессора Попова ул., 5, Санкт-Петербург, 197376, РФ

**Введение:** решение задачи согласованности данных и процессов в условиях децентрализованной асинхронной среды является актуальной для распределенных систем, таких как глобальные облачные платформы, интернет вещей, блокчейн-инфраструктура. Существующие протоколы синхронизации основаны на предположении о симметричности временной задержки на передаче и приеме, что не соответствует действительности в условиях меняющейся сетевой нагрузки, а стохастическая природа сетевых шумов игнорируется. Указанные недостатки протоколов приводят к некорректной синхронизации времени. **Цель:** разработать гибридный метод синхронизации времени, основанный на фильтре Калмана для сглаживания сетевых шумов и логических часах для адаптации к меняющейся сетевой нагрузке. **Результаты:** показано, что существующие протоколы синхронизации распределенных систем становятся ненадежными: на синхронизацию времени влияют сетевые задержки, джиттер и аппаратные погрешности. Приведены ограничения современных подходов синхронизации времени: в работе протокола NTP синхронизации клиента и сервера не учитывается асимметрия каналов; в работе протокола RTP синхронизации локальных сетей фиксированные интервалы синхронизации приводят к шуму и накоплению ошибок; гибридизация протоколов NTP и RTP не обеспечивает динамическую адаптацию к изменяющимся условиям. Предложена гибридная модель синхронизации времени на базе комбинации фильтра Калмана и логических часов. Фильтр Калмана позволяет эффективно подавлять сетевой джиттер и компенсировать дрейф физических часов, а логические часы позволяют быстрее адаптироваться к меняющейся сетевой нагрузке. **Практическая значимость:** внедрение метода на базе комбинации фильтра Калмана и логических часов может стать экономичной альтернативой специализированным протоколам синхронизации в средах с нестабильной нагрузкой.

**Ключевые слова** – синхронизация времени, распределенные системы, физические часы, логические часы, источник точного времени, меняющаяся сетевая нагрузка, фильтр Калмана.

**Для цитирования:** Татарникова Т. М., Архипцев Е. Д. Гибридный метод синхронизации времени в распределенных системах. *Информационно-управляющие системы*, 2025, № 4, с. 26–35. doi:10.31799/1684-8853-2025-4-26-35, EDN: XAGEEQ

**For citation:** Tatarnikova T. M., Arkhiptsev E. D. Hybrid method of time synchronization in distributed systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 26–35 (In Russian). doi:10.31799/1684-8853-2025-4-26-35, EDN: XAGEEQ

### Введение

В распределенных системах, таких как глобальные облачные платформы, интернет вещей (IoT), блокчейн-инфраструктура, остается актуальной проблема согласованности процессов в условиях децентрализованной асинхронной среды [1–3].

Одной из основных задач в этом направлении является синхронизация времени, и чем выше сетевая нагрузка, тем сложнее поддерживать решение данной задачи. В теории методы синхронизации для распределенных систем должны одновременно обеспечивать следующие характеристики: метрическую точность, каузальную согласованность и устойчивость к динамическим сетевым условиям [4, 5].

Метрическая точность – это точность измерения физического времени на всех узлах распределенной системы.

Каузальная согласованность – это модель согласованности, которая отражает потенциальные причинно-следственные связи между операциями и гарантирует, что все процессы соблюдают их в общем порядке [6]. Например, в мессенджерах каузальная согласованность позволяет избежать неверного отображения сообщений и изменений, т. е. случаев, когда ответ на сообщение может появиться раньше оригинального, если узлы не синхронизированы по времени.

Устойчивость к динамическим сетевым условиям – это способность сети возвращаться в свое синхронное состояние после того, как она столкнулась с помехами.

Однако на практике, подобно теореме CAP (Consistency – согласованность, Availability – доступность, Partition tolerance – устойчивость к фрагментации), устанавливающей, что в лю-

бой реализации распределенных не транзакционных баз данных возможно обеспечить не более двух из трех свойств [7], для реализации синхронизации в распределенных системах разработчикам протоколов также необходимо определять приоритеты и находить баланс между требуемыми характеристиками.

Синхронизация в распределенных системах основана на таких протоколах, как NTP (Network Time Protocol) – синхронизация клиента и сервера и PTP (Precision Time Protocol) – синхронизация в локальных сетях [8]. С появлением распределенных систем обработки больших данных на синхронизацию времени стали влиять сетевые задержки, джиттер и аппаратные погрешности [9, 10]. Соответственно, применение этих протоколов стало ненадежным в сценариях с жесткими требованиями к синхронизации времени.

Попытки гибридизации подходов к синхронизации времени по протоколам NTP и PTP обсуждались в работах [11, 12], по результатам которых можно сделать вывод о потенциале комбинирования физических и логических меток времени, например механизме HLC (Hybrid Logical Clocks), сочетающем физическое время с логическими счетчиками. Однако предлагаемые решения не обеспечивают динамическую адаптацию к изменяющимся условиям, к тому же стохастическая природа сетевых шумов игнорируется [13].

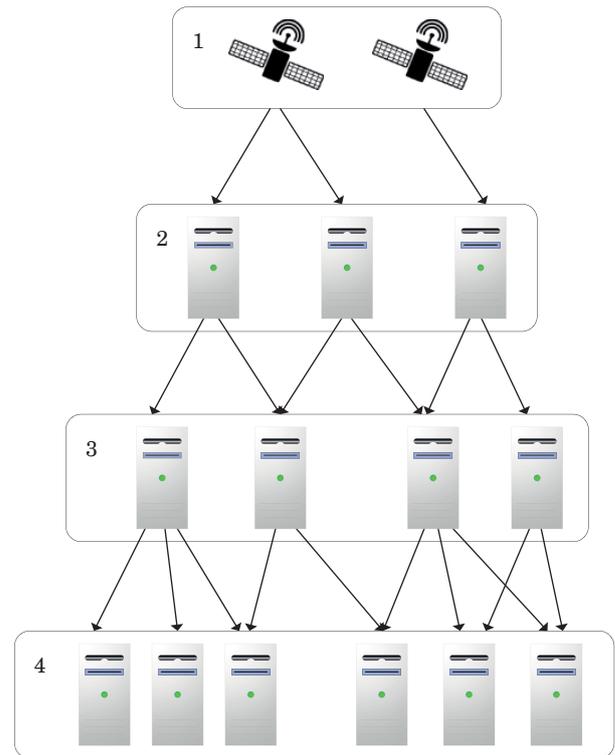
В статье предлагается гибридный метод синхронизации времени – фильтр Калмана в интеграции с логическими часами.

Известно, что фильтр Калмана успешно применяется в навигационных системах для оценки состояния через зашумленные измерения. Поэтому его интеграция с логическими часами может стать основой для создания адаптивного механизма, который не только корректирует физическое время, учитывая статистические аномалии, но и сохраняет причинно-следственную целостность событий.

### Анализ ограничений современных подходов синхронизации времени

Network Time Protocol – наиболее распространенный протокол синхронизации времени, основанный на иерархии серверов точного времени Stratum (рис. 1).

Чем выше уровень сервера Stratum, тем он ближе к эталонному источнику и тем выше его метрическая точность. Серверы верхнего уровня (Stratum 1) получают время напрямую от атомных часов или GPS, используя аппаратное соединение через спутник, телефонные модемы



■ Рис. 1. Иерархия Stratum серверов  
■ Fig. 1. Hierarchy of Stratum servers

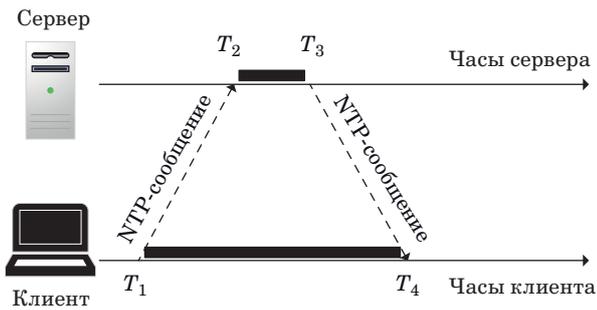
или радио. Для них характерны задержки до 1 мс при использовании локального источника – встроенных GPS-приемников и до 10 мс при использовании удаленного источника – радио или интернета. Доступ к Stratum 1 является ограниченным и предоставляется доверенным клиентам или в рамках исследовательских или представительских сетей, поскольку они имеют ограниченную пропускную способность. Серверы нижних уровней (Stratum 2, Stratum 3 и т. д.) в основном являются публичными и используют сетевое соединение. Серверы уровня Stratum 2 опрашивают несколько серверов уровня Stratum 1 для повышения точности и имеют погрешность до 50 мс [14].

В работе протокола NTP выделяются четыре основные временные точки (рис. 2):  $T_1$  – время отправки запроса от клиента к серверу (по часам клиента);  $T_2$  – время получения запроса сервером (по часам сервера);  $T_3$  – время отправки ответа сервером (по часам сервера);  $T_4$  – время получения ответа клиентом (по часам клиента).

Задержка *Delay* и смещение *Offset* вычисляются следующим образом:

$$Delay = (T_4 - T_1) - (T_3 - T_2); \quad (1)$$

$$Offset = \frac{(T_2 - T_1) - (T_4 - T_3)}{2}, \quad (2)$$



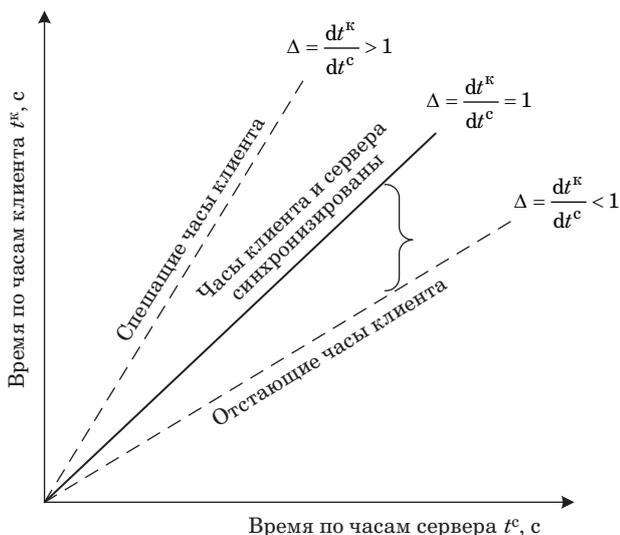
■ **Рис. 2.** Синхронизация времени между клиентом и сервером

■ **Fig. 2.** Time synchronization between client and server

где продолжительность интервалов  $(T_2 - T_1)$  и  $(T_4 - T_3)$  включает время распространения NTP-сообщения и разницу  $\delta$  в показаниях часов клиента и сервера. На рис. 3 демонстрируются соотношения времени по часам клиента и часам сервера: рассинхронизация и дрейф частоты.

Для корректировки собственных часов клиенту необходимо добавить вычисленное значение *Offset* к текущему времени:  $NewTime = CurrentTime + Offset$ .

Работа протокола NTP основана на предположении о том, что сетевые задержки симметричны – имеют нормальное распределение. Если путь «Клиент → Сервер» и «Сервер → Клиент» имеют разную задержку (асимметрия каналов), то протокол NTP определит некорректное значение смещения. На практике так и получается. Например, в средах с беспроводным доступом Wi-Fi или сотовой связи задержки группируются вокруг нескольких типичных значений из-за



■ **Рис. 3.** Соотношение времени по часам клиента и сервера

■ **Fig. 3.** Time ratio between client and server clocks

нестабильности среды. Это связано с тем, что пакеты сериализуются, ожидают в очередях роутеров, тем самым создаются тяжелые хвосты в распределении сетевой задержки. Таким образом, работа протокола NTP в сетях подобного типа систематически приводит к ошибочной оценке *Offset* [15].

В протоколе PTP, известного как стандарт IEEE 1588, более высокая метрическая точность, чем в протоколе NTP, достигается за счет обязательного использования специализированного оборудования, способного на физическом уровне взаимодействия PHY (Physical Layer Protocol) предоставлять метки времени в передаваемых пакетах времени непосредственно перед моментом передачи их в канал связи. Это позволяет минимизировать джиттер, вносимый операционными системами и стеками протоколов, а также обеспечивает автоматическое восстановление при отказах.

В распределенной системе, синхронизированной по протоколу PTP, различают следующие типы узлов: гроссмейстерские часы (Grandmaster) – основной источник, по которому синхронизируется время в системе; ведущие часы – источники точного времени для конечных узлов; ведомые часы – конечные узлы, на которых необходимо осуществить синхронизацию времени по протоколу PTP.

Выбор Grandmaster осуществляется динамически на основе специального механизма Best Master Clock Algorithm (BMCA). При потере узлом Grandmaster связи с физическими часами его место автоматически занимает следующий узел, претендующий на роль Grandmaster, который становится источником точного времени для конечных узлов. С этой целью все ведущие часы системы постоянно находятся в режиме прослушивания сообщений, отправленных на широковещательный адрес протокола PTP.

Основными недостатками протокола PTP являются предположение о симметричности задержек в направлениях «Ведущий → Ведомый» и «Ведомый → Ведущий» и фиксированный интервал синхронизации. Правда, в последней версии стандарта в протокол введены компенсации асимметричных задержек [16, 17] через калибровку смещений, но для этого необходимо использовать сторонние протоколы, оборудование и т. д. Фиксированные интервалы синхронизации в сценариях с резким изменением сетевой нагрузки, например flash-трафик в дата-центре, приводят к «раскачиванию» часов: частые коррекции вводят высокочастотный шум, а редкие вызывают накопление ошибок. Также немаловажным является отсутствие встроенной криптографической защиты в протоколе PTP, что дает возможность злоумышленнику искусствен-

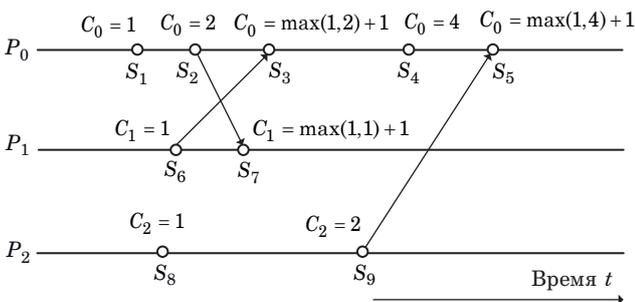
но увеличивать время доставки пакетов или фальсифицировать сообщения для назначения ложного Grandmaster. Тем не менее наличие механизма компенсации задержек является основным преимуществом протокола РТР, что делает его незаменимым в промышленных системах управления, телекоммуникациях и высокочастотном трейдинге, где погрешность синхронизации свыше 1 мс неприемлема [18].

Механизм НЛС синхронизации времени в распределенных системах сочетает физическое время с логическими часами Лэмпорта и векторными часами. Механизм обеспечивает не только метрическую точность, но и частично каузальную согласованность.

Логические часы Лэмпорта представляют собой счетчик определения порядка событий. Работа счетчика описывается следующей схемой (рис. 4):

- монотонное увеличение счетчика  $C$  при выполнении внутренних событий  $S$  каждого процесса  $P$ ;
- значение счетчика  $i$ -го процесса передается  $j$ -му процессу в пакете времени;
- при получении  $j$ -м процессом пакета времени значение счетчика  $j$ -го процесса устанавливается как максимальное из текущего значения собственного счетчика и полученного значения счетчика  $i$ -го процесса и увеличивается на 1.

Порядок изменения событий подчиняется правилу: если событие  $S_1$  произошло до  $S_2$ , то временная метка  $S_1$  меньше  $S_2$ :  $TimeStamp(S_1) < TimeStamp(S_2)$ . Однако обратное утверждение неверно, поскольку для событий без детерминистской связи нельзя выстроить временную зависимость. Например, если событие  $S_1$  в процессе  $P_i$  отправляет сообщение, а в процессе  $P_j$  событие  $S_2$  получает это сообщение, то можно сказать, что событие  $S_1$  произошло раньше события  $S_2$ . Предположим, что до того, как процесс  $P_j$  получил сообщение о событии  $S_2$ , в  $P_j$  произошло событие  $S_3$ , т. е.  $S_3$  произошло раньше  $S_2$ , и было отправлено сообщение процессу  $P_i$ , а затем процесс  $P_i$  получил сообщение



■ **Рис. 4.** Часы Лэмпорта  
 ■ **Fig. 4.** Lamport clock

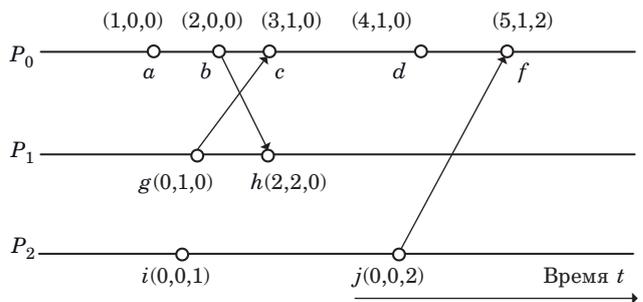
о событии  $S_4$  после события  $S_1$ , т. е. событие  $S_3$ , как и событие  $S_1$ , произошло раньше события  $S_4$ . Основываясь на алгоритме Лэмпорта, невозможно определить какие-либо связи между событиями  $S_1$  и  $S_3$ . Все, что известно, это  $S_1 \rightarrow S_2$ , и  $S_3 \rightarrow S_4$ , и  $S_1 \rightarrow S_4$ , и  $S_3 \rightarrow S_2$ , но непонятно,  $S_1 \rightarrow S_3$  или  $S_3 \rightarrow S_1$ . Таким образом, логические часы Лэмпорта способны обеспечить только частичный порядок, т. е. они указывают на причинно-следственную связь между некоторыми, но не всеми событиями. Это может привести к неопределенности при определении точного порядка событий.

Векторные часы – это расширение возможностей логических часов Лэмпорта. Каждый процесс  $P_i$ ,  $i = 1, \dots, N$  в системе поддерживает вектор логических часов  $(1, \dots, i, \dots, N)$ , где  $i$ -й элемент вектора представляет состояние часов  $i$ -го активного процесса. Приведенный на рис. 5 пример иллюстрирует механизм векторных часов для трех процессов.

Работа векторных часов описывается следующей схемой:

- все часы всегда установлены на ноль;
- каждый раз, когда в  $i$ -м процессе происходит внутреннее событие, значение логических часов  $i$ -го процесса в векторе увеличивается на 1;
- каждый раз, когда  $i$ -й процесс отправляет сообщение  $j$ -му процессу, значение логических часов  $i$ -го процесса в векторе увеличивается на 1;
- каждый раз, когда  $j$ -й процесс получает сообщение, он увеличивает значение логических часов  $i$ -го процесса в векторе на 1;
- каждый элемент вектора часов обновляется путем взятия максимального значения как из собственного вектора, так и из вектора в полученном сообщении для каждого элемента.

Таким образом, сравнивая векторные часы, система может определить, произошло ли событие на одном узле до, после или одновременно с событием на другом узле, что позволяет эффективно разрешать конфликты и обеспечивать каузальную согласованность [19, 20].



■ **Рис. 5.** Векторные часы  
 ■ **Fig. 5.** Vector clock

Основными ограничениями векторных часов, которые влияют на их применимость в современных распределенных системах, являются [21]:

– требование хранения  $N$  часов в системе с  $N$  узлами, что приводит к высоким затратам на коммуникацию;

– сложная реализация векторных часов в сценариях, когда узлы часто присоединяются и покидают систему или при частых сбоях сети;

– увеличение объема передаваемых сообщений из-за необходимости периодического обмена векторными часами между процессами, что становится проблемой для систем с ограниченной пропускной способностью или приложений, чувствительных к задержкам.

При реализации механизма HLC каждый узел в системе поддерживает гибридную метку времени, состоящую из двух компонент:

– физического времени ( $pt$ ) – локального времени узла, синхронизированного с NTP или другим протоколом;

– логического счетчика ( $lc$ ) – счетчика, который необходим для регистрации локальных событий или для разрешения конфликтов.

При получении нового события на узле генерируется новая метка по следующему правилу:

$$\begin{aligned} pt_{new} &= \max(pt_{old}, pt_{local}); \\ lc_{new} &= lc_{old} + 1 \text{ при } pt_{local} > pt_{old}, \end{aligned} \quad (3)$$

где  $pt_{local}$  – текущее физическое время узла, полученное из пакета NTP; ( $pt_{old}, lc_{old}$ ) – текущая метка HLC до обновления.

Перед отправкой сообщения генерируется новая метка HLC, как при локальном событии. На узел отправляется сообщение с меткой ( $pt_{new}, lc_{new}$ ). При получении сообщения на узле вычисляется новая метка:

$$\begin{aligned} pt_{new} &= \max(pt_{current}, pt_{msg}, pt_{local}); \\ lc_{new} &= lc_{current} + 1, lc_{msg} + 1 \text{ при } pt_{msg} = pt_{new}, \end{aligned} \quad (4)$$

где ( $pt_{current}, lc_{current}$ ) – текущая метка HLC на принимающем узле; ( $pt_{msg}, lc_{msg}$ ) – физическая метка, извлекаемая из полученного сообщения  $msg$ .

Таким образом, механизм HLC обеспечивает синхронизацию с физическим временем и каузальную согласованность. Но поскольку механизм HLC основан на физическом времени, которое может быть неточным из-за сбоя NTP или аппаратных ошибок, то метрическая точность не гарантируется.

Как показывает анализ, ни один из существующих протоколов – NTP, RTP, HLC – не решает эту задачу синхронизации в полной мере. Однако комбинация их сильных сторон с адаптивными алгоритмами позволяет создавать гибридные модели, способные компенсировать индивидуаль-

ные недостатки. Одно из таких решений – комбинация фильтра Калмана и логических часов.

### Комбинация фильтра Калмана и логических часов

Для повышения точности синхронизации времени применим фильтр Калмана – адаптивный алгоритм, позволяющий минимизировать шум измерений и предсказывать истинное значение временных меток. Данный метод основан на статистической обработке поступающих измерений и динамическом обновлении оценок с учетом априорной модели изменений времени.

Применение фильтра Калмана позволяет:

– сглаживать вариации задержек передачи данных в NTP, минимизируя ошибки синхронизации;

– корректировать временные метки RTP при резких изменениях мастер-часов;

– обеспечивать устойчивость HLC к неточностям в измерениях физического времени.

Фильтр Калмана опирается на две ключевые модели.

Модель процесса прогнозирования состояния

$$\mathbf{x}_t = \mathbf{F} \times \mathbf{x}_{t-1} + \mathbf{B} \times u_{t-1} + \mathbf{w}_t, \quad (5)$$

где  $\mathbf{x}_t$  – вектор состояния (смещение часов *Offset* и дрейф частоты *Drift*) в текущий момент времени  $t$ ;

$\mathbf{F} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$  – матрица перехода (модель эволюции состояния),  $\Delta t$  – интервал дискретизации;

$\mathbf{B} = \begin{bmatrix} \Delta t \\ 1 \end{bmatrix}$  – матрица-столбец применения управляющего воздействия;

$u_{t-1}$  – управляющее воздействие, по сути значение *Offset* – коррекция от NTP в предыдущий момент времени ( $t - 1$ );  $\mathbf{w}_t$  – вектор шума процесса (белый гауссов шум) в текущий момент времени  $t$ .

Модель наблюдения, которая связывает состояние с измерениями:

$$\mathbf{z}_t = \mathbf{H} \times \mathbf{x}_t + \mathbf{v}_t, \quad (6)$$

где  $\mathbf{z}_t$  – измерение (разница между показаниями часов клиента и сервера);  $\mathbf{H}$  – матрица наблюдений;  $\mathbf{x}_t$  – вектор состояния в текущий момент времени  $t$ ;  $\mathbf{v}_t$  – шум измерений (джиттер задержки *Delay*).

Шум процесса представляется ковариационной матрицей

$$\mathbf{Q}_t = \begin{bmatrix} \mathbf{w}_t \times \mathbf{w}_t^T \end{bmatrix} = \begin{bmatrix} \sigma_{Offset}^2 & 0 \\ 0 & \sigma_{Drift}^2 \end{bmatrix}, \quad (7)$$

где  $\sigma_{Offset}^2$ ,  $\sigma_{Drift}^2$  – дисперсии шумов смещения и дрейфа частоты.

Шум измерений представляется ковариационной матрицей

$$\mathbf{R}_t = \begin{bmatrix} \mathbf{v}_t \times \mathbf{v}_t^T \end{bmatrix}. \quad (8)$$

Фильтр Калмана реализует итеративный процесс, состоящий из двух основных этапов: прогноза – предсказания состояния системы (5) на основе модели наблюдения (6) и коррекции – уточнения предсказания с использованием новых измерений. Эти этапы образуют замкнутый цикл, позволяющий адаптивно балансировать между доверием к модели и поступающим данным.

Для коррекции предсказания применяются следующие выражения:

$$\mathbf{K}_t = \mathbf{P}_t \times \mathbf{H}^T \times (\mathbf{H} \times \mathbf{P}_t \times \mathbf{H}^T + \mathbf{R}_t)^{-1}; \quad (9)$$

$$\mathbf{x}_t = \mathbf{x}_t^- + \mathbf{K}_t \times (\mathbf{z}_t - \mathbf{H} \times \mathbf{x}_t^-); \quad (10)$$

$$\mathbf{P}_t = \left[ (\mathbf{x}_t - \mathbf{x}_t^-) \times (\mathbf{x}_t - \mathbf{x}_t^-)^T \right], \quad (11)$$

где  $\mathbf{K}_t$  – коэффициент Калмана (коэффициент усиления фильтра);  $\mathbf{P}_t$  – матрица ковариации предсказываемой ошибки оценивания;  $\mathbf{x}_t^-$  – вектор предсказываемой оценки состояния;  $\mathbf{P}_t$  – матрица ковариации ошибки оценивания.

Алгоритм работы фильтра Калмана.

Шаг 1. Инициализация фильтра Калмана:  $\mathbf{x}_0$  – начального значения вектора предсказываемой оценки состояния и  $\mathbf{P}_0$  – матрицы ковариации предсказываемой ошибки оценивания.

Шаг 2. Измерение  $\mathbf{z}_t$ .

Шаг 3. Вычисление коэффициента Калмана по формуле (9).

Шаг 4. Коррекция вектора состояний по формуле (10).

Шаг 5. Коррекция матрицы ковариации ошибки оценивания по формуле (11).

Шаг 6. Предсказание следующего шага итерации:

$$\mathbf{x}_t^- = \mathbf{F} \times \mathbf{x}_{t-1} + \mathbf{B} \times \mathbf{u}_{t-1};$$

$$\mathbf{P}_t = \mathbf{F} \times \mathbf{P}_{t-1} \times \mathbf{F}^T + \mathbf{Q}.$$

Шаг 7. Переход на шаг 2 с повторением процедуры фильтрации Калмана для каждого нового измерения.

## Результаты натурального эксперимента

Для эксперимента было создано четыре виртуальные вычислительные машины (ВВМ) с соответствующими методами синхронизации

времени: по протоколу NTP, по протоколу RTP, HLC с часами Лэмпорта и HLC с векторными часами. На каждом сервере генерировалась нагрузка в виде постоянных HTTP-запросов и применялся гибридный метод интеграции фильтра Калмана с логическими часами. Параметры фильтра Калмана  $\mathbf{Q}$  подбирались экспериментальным путем:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0,1 & 0 \\ 0 & 0,01 \end{bmatrix}. \quad (12)$$

Замеры *Offset* на ВВМ#1 выполнены с синхронизацией времени по протоколу NTP, так называемый сырой *Offset*, и с синхронизацией гибридным методом интеграции фильтра Калмана с логическими часами (рис. 6, а). Гистограмма изменения джиттера задержки *Delay* на том же самом интервале замеров представлена на рис. 6, б.

Продолжительность эксперимента составила 6 ч, но на графиках приведены замеры на 100 с, отражающие наиболее общую картину результатов эксперимента.

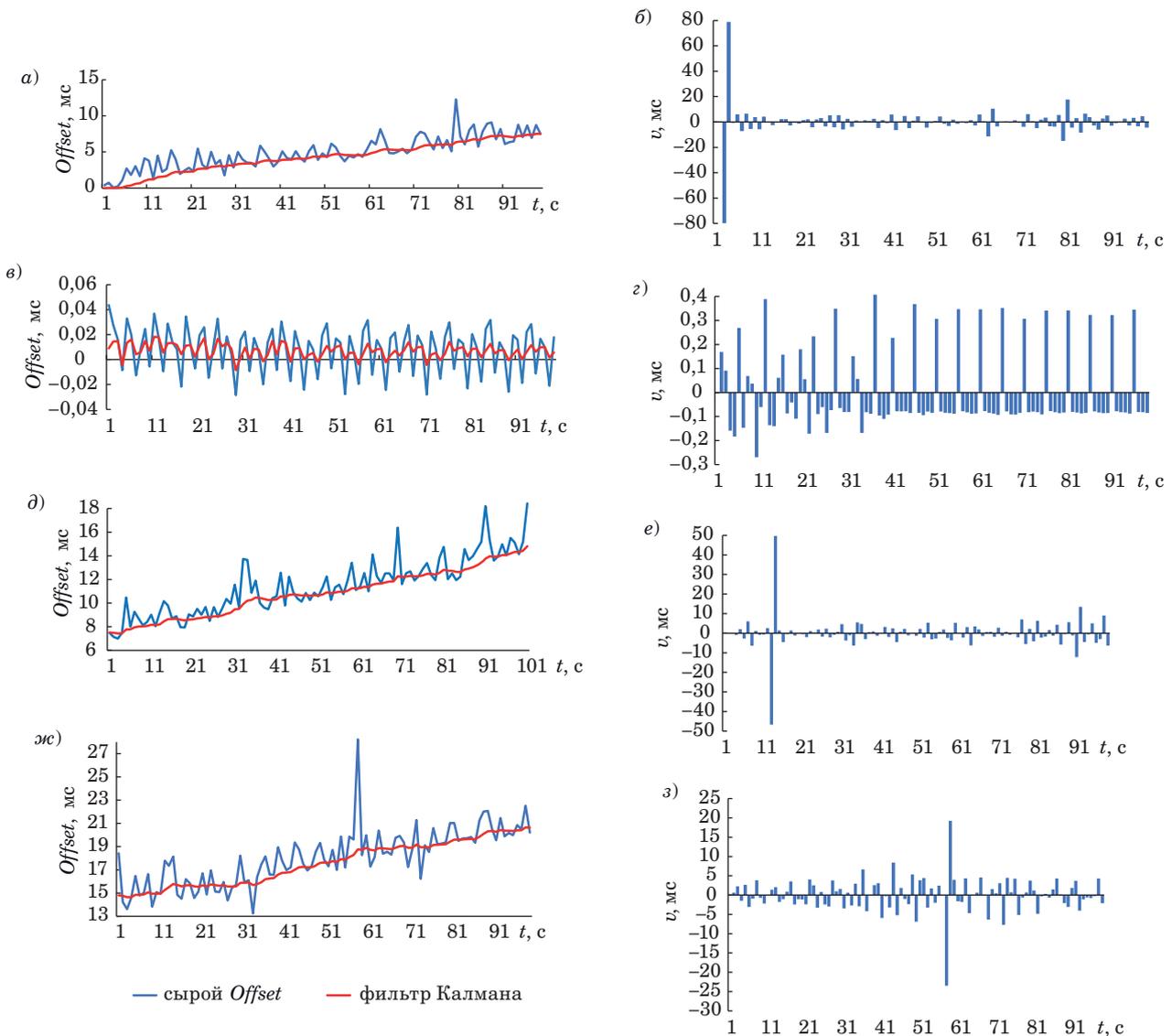
Без использования фильтра Калмана среднеквадратическое отклонение джиттера составило 12,23 мс, с фильтром Калмана оно стремится к нулю – по этой причине приводить гистограмму изменения джиттера после фильтра Калмана не имеет смысла. Таким образом, при использовании гибридного метода интеграции фильтра Калмана с логическими часами удалось снизить среднеквадратическое отклонение джиттера на 74 %, а максимальный *Offset* с 12,3 до 7,5 мс.

С синхронизацией времени по протоколу RTP и с синхронизацией гибридным методом интеграции фильтра Калмана с логическими часами проведены замеры *Offset* на ВВМ#2 (рис. 6, в). Гистограмма изменения джиттера на том же самом интервале замеров представлена на рис. 6, з.

Без использования фильтра Калмана среднеквадратическое отклонение джиттера по протоколу RTP составило 0,17 мс, с фильтром Калмана оно стремится к нулю. При использовании гибридного метода интеграции фильтра Калмана с логическими часами удалось снизить среднеквадратическое отклонение джиттера на 58 %, а максимальный *Offset* – с 0,044 до 0,018 мс.

Результаты замеров *Offset* на ВВМ#3 с синхронизацией времени по механизму HLC с часами Лэмпорта и с синхронизацией гибридным методом интеграции фильтра Калмана с логическими часами показаны на рис. 6, д. Гистограмма изменения джиттера на том же самом интервале замеров представлена на рис. 6, е.

Без использования фильтра Калмана среднеквадратическое отклонение джиттера составило 7,79 мс, с фильтром Калмана оно стремится к ну-



■ **Рис. 6.** Графики замеров *Offset* (слева) и гистограмма изменения джиттера (справа) на ВВМ#1 (а, б); ВВМ#2 (в, з); ВВМ#3 (д, е); ВВМ#4 (ж, з)

■ **Fig. 6.** *Offset* measurement graphs (left) and jitter change histogram (right) on virtual machine #1 (а, б); virtual machine #2 (в, з); virtual machine #3 (д, е); virtual machine #4 (ж, з)

лю. При использовании гибридного метода интеграции фильтра Калмана с логическими часами удалось снизить среднеквадратическое отклонение джиттера на 52 %, а максимальный *Offset* — с 18,43 до 14,8 мс.

Графики замеров *Offset* на ВВМ#4 с синхронизацией времени по механизму NLC с векторными часами и с синхронизацией гибридным методом интеграции фильтра Калмана с логическими часами приведены на рис. 6, ж. Гистограмма изменения джиттера на том же самом интервале замеров представлена на рис. 6, з.

Без использования фильтра Калмана среднеквадратическое отклонение джиттера составило 4,36 мс, с фильтром Калмана оно стремится к ну-

лю. При использовании гибридного метода интеграции фильтра Калмана с логическими часами удалось снизить среднеквадратическое отклонение джиттера на 47 %, а максимальный *Offset* — с 28,23 до 20,68 мс.

### Закключение

Интеграция фильтра Калмана в классические протоколы синхронизации времени, такие как NTP или PTP, обеспечивает существенное повышение точности синхронизации часов клиента и сервера в распределенных системах за счет принципиально иного подхода к обработке

неопределенностей, присущих сетевым взаимодействиям. Ключевое преимущество заключается в способности алгоритма Калмана декомпонировать системные шумы и полезный сигнал на основе стохастической модели, что невозможно в рамках детерминированных методов коррекции, используемых стандартными протоколами.

Таким образом, гибридный метод, сочетающий применение фильтра Калмана с логическими часами, преодолевает фундаментальное ограничение протоколов синхронизации – зависимость точности от стационарности сетевой среды. Это возможно благодаря наличию адаптивной стохастической обратной связи – коррекции состояния системы на основе модели наблюдения.

Эксперимент подтверждает, что фильтр Калмана – эффективный инструмент для борьбы с джиттером в распределенных системах, особенно в NTP-сетях. При помощи него удалось снизить среднеквадратическое отклонение джиттера и *Offset* для всех рассмотренных протоколов и механизмов синхронизации времени.

Внедрение метода в облачные инфраструктуры может стать решением для задач, где важна балансировка между точностью и каузальной согласованностью.

Дальнейшие исследования целесообразно направить на создание самообучающихся алгоритмов для динамических сред.

## Литература

- Balakrishnan K., Dhanalakshmi R., Vam Bahadur Sinha, Gopalakrishnan R.** Clock synchronization in industrial Internet of Things and potential works in precision time protocol: Review, challenges and future directions. *International Journal of Cognitive Computing in Engineering*, 2023, vol. 4, pp. 205–219. <https://doi.org/10.1016/j.ijcce.2023.06.001>
- Vilaplana J., Solsona F., Teixidó I., Mateo Fornes J.** A queuing theory model for cloud computing. *The Journal of Supercomputing*, 2014, vol. 69, pp. 492–507. doi:10.1007/s11227-014-1177-y
- Özsu M. T., Valduriez P.** *Principles of Distributed Database Systems*. Springer Cham, 2020. 676 p.
- Поляков В. Б., Аксёнов Н. В., Сергеев М. В.** Средства обеспечения временной синхронизации данных и процессов в системах интернет вещей на уровне устройств и их классификация. *Альманах современной метрологии*, 2019, № 4 (20), с. 40–53. EDN: PFXPMY
- Tatarnikova T. M., Arkhiptsev E. M.** Designing fault-tolerant systems with micro-service architecture. *XXVII International Conference on Soft Computing and Measurements (SCM)*, Saint-Petersburg, Russian Federation, 2024, pp. 348–351. doi:10.1109/SCM62608.2024.10554143
- Elbushra M. M., Lindström J.** Causal consistent databases. *Open Journal of Databases*, 2015, vol. 2, no. 1, pp. 17–35.
- Татарникова Т. М., Архипцев Е. Д., Кармановский Н. С.** Определение размера кластера и числа реплик высоконагруженных информационных систем. *Известия высших учебных заведений. Приборостроение*, 2023, т. 66, № 8, с. 646–651. doi:10.17586/0021-3454-2023-66-8-646-651
- Рыжков А. В., Колтунов М. Н., Насонов А. Ю., Шварц М. Л.** Проблемы сличения шкал времени в пакетных сетях электросвязи. *T-Comm: Телекоммуникации и транспорт*, 2017, т. 11, № 11, с. 10–17.
- Фунг В. К., Богатырев В. А.** Задержки и надежность обслуживания запросов в виртуальном компьютерном кластере. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*, 2024, № 68, с. 48–58. doi:10.17223/19988605/68/5
- Воронин В. В., Землянухина А. В., Репецкий А. А.** Уязвимость NTP-протокола: анализ происшествий и стратегия безопасности. *Вестник ТОГУ*, 2025, № 1(76), с. 49–58. <https://doi.org/10.38161/1996-3440-2025-1-49-58>
- Mallela C., Tholu K., Bordogna M.** Timing models for PTP in Ethernet networks. *2017 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*, 2017, pp. 1–6.
- Kulkarni S. S., Demirbas M., Madeppa D., Avva B., Leone M.** Logical physical clocks. *Principles of Distributed Systems: 18th International Conference*, 2014, pp. 17–32.
- Zhou K., Wang X., Cui S.** Time synchronization in Industrial Internet of Things: A survey. *IEEE 7th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2024, vol. 7, pp. 506–510.
- Rinaldi S., Giustina D. D., Ferrari P., Flammini A., Sisinni E.** Time synchronization over heterogeneous network for smart grid application: Design and characterization of a real case. *Ad Hoc Networks*, 2016, vol. 50, pp. 41–57. <https://doi.org/10.1016/j.adhoc.2016.04.001>
- Губанов Н. Г., Рагузин А. С.** Анализ и моделирование протоколов синхронизации в информационной системе региональной электросетевой компании. *Вестник Самарского государственного технического университета. Серия: Технические науки*, 2016, № 1(49), с. 7–14.
- Exel R.** Mitigation of asymmetric link delays in IEEE 1588 clock synchronization systems. *IEEE Communications Letters*, 2014, vol. 18, no. 3, pp. 507–510.

17. Батенков К. А., Цицин Е. А. Подстройка параметров времени для сетевых устройств с помощью протокола синхронизации РТР (Precision Time Protocol). *Инфокоммуникационные технологии: актуальные вопросы цифровой экономики: сборник научных трудов I Международной научно-практической конференции*, Екатеринбург, 17–18 февраля 2021 г. Екатеринбург, УрТИСИ, 2021, с. 90–95.
18. Мазуренко Д. К. Сетевые решения построения пакетной сети распределения сигналов единого точного времени. *T-Comm: Телекоммуникации и транспорт*, 2015, т. 9, № 4, с. 67–71.
19. Finkenzeller A., Butowski O., Regnath E., Hamad M., Steinhorst S. PTPsec: Securing the Preci-

sion Time Protocol against time delay attacks using cyclic path asymmetry analysis. *IEEE INFOCOM 2024 – IEEE Conference on Computer Communications*, 2024, pp. 461–470.

20. Bogatyrev V. A., Derkach A. N., Aleksankov S. M. The model of reliability of dulated real-time computers for cyber-physical systems. *Studies in Systems, Decision and Control*, 2020, vol. 260, pp. 11–21. doi:10.1007/978-3-030-32648-7\_2, EDN: ZXAHKJ
21. Tatarnikova T. M., Arkhiptsev E. D. Determine the number of distributed big data storage replicas. *XXVI International Conference on Soft Computing and Measurements (SCM)*, 2023, pp. 223–226. doi:10.1109/SCM58628.2023.10159087

UDC 004.032.26

doi:10.31799/1684-8853-2025-4-26-35

EDN: XAGEEQ

### Hybrid method of time synchronization in distributed systems

T. M. Tatarnikova<sup>a</sup>, Dr. Sc., Tech., Professor, orcid.org/0000-0002-6419-0072, tm-tatarn@yandex.ru

E. D. Arkhiptsev<sup>b</sup>, Post-Graduate Student, orcid.org/0009-0006-3001-953X

<sup>a</sup>Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

<sup>b</sup>Saint-Petersburg Electrotechnical University «LETI», 5, Prof. Popov St., 197376, Saint-Petersburg, Russian Federation

**Introduction:** The solution to the problem of data and process consistency in a decentralized asynchronous environment is relevant for distributed systems such as global cloud platforms, the Internet of Things, and blockchain infrastructure. Existing synchronization protocols are based on the assumption of symmetrical time delays during transmission and reception, which is not true in the conditions of changing network load in addition, the stochastic nature of network noise is ignored. These shortcomings of the protocols lead to incorrect time synchronization. **Purpose:** To develop a hybrid time synchronization method based on the Kalman filter to smooth out network noise and on a logical clock for the system to adapt to the changing network load. **Results:** We demonstrate that the existing distributed systems synchronization protocols are becoming unreliable as time synchronization is affected by network delays, jitter, and hardware errors. We highlight the limitations of modern time synchronization: the NTP protocol for synchronizing Clients and Servers does not consider channel asymmetry; fixed synchronization intervals in the PTP protocol for synchronizing local networks lead to noise and error accumulation; hybridization of the NTP and PTP protocols does not provide dynamic adaptation to changing conditions. We propose a hybrid time synchronization model based on the combination of the Kalman filter and a logical clock. The Kalman filter makes it possible to effectively suppress network jitter and to compensate for physical clock drift, while logical clocks enable a faster adaptation to changing network loads. The results of a full-scale experiment demonstrate a more than threefold reduction in average delay. **Practical relevance:** The implementation of the method based on the combination of the Kalman filter and a logical clock can become a cost-effective alternative to specialized synchronization protocols in environments with unstable loads.

**Keywords** – time synchronization, distributed systems, physical clock, logical clock, precise time source, changing load, Kalman filter.

**For citation:** Tatarnikova T. M., Arkhiptsev E. D. Hybrid method of time synchronization in distributed systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 26–35 (In Russian). doi:10.31799/1684-8853-2025-4-26-35, EDN: XAGEEQ

### References

- Balakrishnan K., Dhanalakshmi R., Bam Bahadur Sinha, Gopalakrishnan R. Clock synchronization in industrial Internet of Things and potential works in precision time protocol: Review, challenges and future directions. *International Journal of Cognitive Computing in Engineering*, 2023, vol. 4, pp. 205–219. https://doi.org/10.1016/j.ijcce.2023.06.001
- Vilaplana J., Solsona F., Teixidó I., Mateo Fornes J. A queuing theory model for cloud computing. *The Journal of Supercomputing*, 2014, vol. 69, pp. 492–507. doi:10.1007/s11227-014-1177-y
- Özsu M. T., Valduriez P. *Principles of Distributed Database Systems*. Springer Cham, 2020. 676 p.
- Polyakov V. B., Aksonov N. V., Sergeev M. V. The means of providing time synchronization of data and processes in the Internet of Things at the level of devices and their classification. *Al'manac of Modern Metrology*, 2019, no. 4 (20), pp. 40–53 (In Russian). EDN: PFXPMY
- Tatarnikova T. M., Arkhiptsev E. M. Designing fault-tolerant systems with micro-service architecture. *XXVII International Conference on Soft Computing and Measurements (SCM)*, Saint-Petersburg, Russian Federation, 2024, pp. 348–351 (In Russian). doi:10.1109/SCM62608.2024.10554143
- Elbushra M. M., Lindström J. Causal consistent databases. *Open Journal of Databases*, 2015, vol. 2, no. 1, pp. 17–35.
- Tatarnikova T. M., Arkhiptsev E. D., Karmanovskiy N. S. Determining the cluster size and the number of replicas of highly loaded information systems. *Journal of Instrument Engineering*, 2023, vol. 66, no. 8, pp. 646–651 (In Russian).
- Ryzhkov A. V., Koltunov M. N., Nasonov A. Y., Schwartz M. L. Problems of comparing time scales in packet telecommunication networks. *T-Comm*, 2017, vol. 11, no. 11, pp. 10–17 (In Russian).
- Phung V. Q., Bogatyrev V. A. Delays and reliability of request servicing in a virtual computer cluster. *Tomsk State*

- University Journal of Control and Computer Science*, 2024, vol. 68, pp. 48–58 (In Russian). doi:10.17223/19988605/68/5
10. Voronin V. V., Zemlyanukhina A. V., Repetskiy A. A. NTP protocol vulnerabilities: Incident analysis and security strategy. *Bulletin of Pacific National University*, 2025, no. 1(76), pp. 49–58 (In Russian). <https://doi.org/10.38161/1996-3440-2025-1-49-58>
  11. Mallela C., Tholu K., Bordogna M. Timing models for PTP in Ethernet networks. *2017 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*, 2017, pp. 1–6.
  12. Kulkarni S. S., Demirbas M., Madeppa D., Avva B., Leone M. Logical physical clocks. *Principles of Distributed Systems: 18th International Conference*, 2014, pp. 17–32.
  13. Zhou K., Wang X., Cui S. Time synchronization in industrial Internet of Things: A survey. *IEEE 7th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2024, vol. 7, pp. 506–510.
  14. Rinaldi S., Giustina D. D., Ferrari P., Flammini A., Sisinni E. Time synchronization over heterogeneous network for smart grid application: Design and characterization of a real case. *Ad Hoc Networks*, 2016, vol. 50, pp. 41–57. <https://doi.org/10.1016/j.adhoc.2016.04.001/>
  15. Gubanov N. G., Raguzin A. S. Analysis and modeling of the existing protocols used in time synchronization of the regional power-grid information systems. *Vestnik Samarskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Tekhnicheskie nauki*, 2016, no. 1(49), pp. 7–14 (In Russian).
  16. Exel R. Mitigation of asymmetric link delays in IEEE 1588 clock synchronization systems. *IEEE Communications Letters*, 2014, vol. 18, no. 3, pp. 507–510.
  17. Batenkov K. A., Tsitsin E. A. Adjusting time settings for networked devices with ptp synchronization protocol. *Sbornik nauchnykh trudov I Mezhdunarodnoy nauchno-prakticheskoy konferentsii "Infokommunikatsionnyye tekhnologii: aktual'nyye voprosy tsifrovoy ekonomiki"* [Proc. I Intern. Scient. and Pract. Conf. "Infocommunication technologies: current issues of the digital economy"]. Ekaterinburg, 2021, pp. 90–95 (In Russian).
  18. Mazurenko D. K. Network reference model of the UTC distribution for the packet networks. *T-Comm*, 2015, vol. 9, no. 4, pp. 67–71 (In Russian).
  19. Finkenzeller A., Butowski O., Regnath E., Hamad M., Steinhorst S. PTPsec: Securing the Precision Time Protocol against time delay attacks using cyclic path asymmetry analysis. *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*, 2024, pp. 461–470.
  20. Bogatyrev V. A., Derkach A. N., Aleksankov S. M. The model of reliability of duplicated real-time computers for cyber-physical systems. *Studies in Systems, Decision and Control*, 2020, vol. 260, pp. 11–21. doi:10.1007/978-3-030-32648-7\_2, EDN: ZXAHKJ
  21. Tatarnikova T. M., Arkhptsev E. D. Determine the number of distributed big data storage replicas. *XXVI International Conference on Soft Computing and Measurements (SCM)*, 2023, pp. 223–226. doi:10.1109/SCM58628.2023.10159087



## Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением

А. В. Сацюк<sup>а</sup>, канд. техн. наук, доцент, [orcid.org/0009-0006-7228-8279](https://orcid.org/0009-0006-7228-8279)

Н. В. Володарец<sup>б</sup>, канд. техн. наук, доцент, [orcid.org/0000-0002-8526-4800](https://orcid.org/0000-0002-8526-4800), [volodarets.nikita@yandex.ru](mailto:volodarets.nikita@yandex.ru)

<sup>а</sup>Донецкий институт железнодорожного транспорта, Горная ул., 6, Донецк, Донецкая Народная Республика, 283018, РФ

<sup>б</sup>Донецкая академия транспорта, Дзержинского пр., 7, Донецк, Донецкая Народная Республика, 283086, РФ

**Введение:** современные системы компьютерного зрения для БПЛА сталкиваются с проблемой надежной детекции и трекинга в реальном времени при ограниченных ресурсах embedded-платформ, особенно при интеграции нейросетевых детекторов с алгоритмами сопровождения. Существующие реализации YOLO, несмотря на популярность, имеют недостатки: избыточную вычислительную сложность из-за ориентации на многоклассовую детекцию и неоптимальное взаимодействие с трекинг-алгоритмами. **Цель:** разработать оптимизированную версию YOLOv8 для гибридной системы детекции и трекинга на борту БПЛА с автоматическим наведением, направленную на снижение вычислительной сложности при сохранении точности и адаптацию к ресурсоограниченным платформам. **Результаты:** исследование основано на экспериментах с модифицированной YOLOv8m, оцененной на embedded-платформе (Raspberry Pi 5) и специализированном датасете ONE\_OBJECT. Разработана модифицированная версия YOLOv8m с избирательной заменой стандартных сверточных слоев на depthwise separable convolutions в блоках СЗСА и части слоев Neck. В результате экспериментальных исследований достигнуто снижение вычислительной сложности на 32,9 % (с 8,5 до 5,7 GFLOPS), количества параметров на 37,1 % (с 25,9 до 16,3 млн) и требований к памяти на 29,4 % (с 102 до 72 МБ). Скорость обработки на Raspberry Pi 5 увеличилась на 63,6 % (с 11 до 18 FPS) при сохранении высокой точности детекции mAP@0.5 на уровне 93,5 % – падение всего на 0,7 процентного пункта относительно базовой модели при тестировании на датасете ONE\_OBJECT. Наибольшее снижение точности (1,2 процентного пункта) наблюдается для объектов малого размера (<50 пикселей). **Практическая значимость:** разработанная модификация YOLOv8m успешно интегрирована с алгоритмом трекинга CSRT. Следовательно, появляется возможность создавать эффективные гибридные системы автоматического наведения для БПЛА. Предложенное решение особенно перспективно для embedded-систем с ограниченными ресурсами. Более того, полученные результаты открывают новые возможности для создания энергоэффективных систем компьютерного зрения реального времени.

**Ключевые слова** – YOLOv8, depthwise separable convolutions, БПЛА, гибридная детекция, трекинг CSRT, модификация, оптимизация нейросетей.

**Для цитирования:** Сацюк А. В., Володарец Н. В. Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением. *Информационно-управляющие системы*, 2025, № 4, с. 36–44. doi:10.31799/1684-8853-2025-4-36-44, EDN: YKQVJU

**For citation:** Satsiuk A. V., Volodarets N. V. Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 36–44 (In Russian). doi:10.31799/1684-8853-2025-4-36-44, EDN: YKQVJU

### Введение

Сегодня активно развиваются интеллектуальные системы во многих отраслях. Так, на транспорте многочисленные исследования посвящены вопросам управления транспортными средствами, включая автоматическое регулирование движения транспортных средств [1, 2], разработку систем автономного управления [3–7], а также оптимизацию параметров, влияющих на движение [8–12]. Особого подхода в решении данных вопросов требуют системы управления беспилотными летательными аппаратами (БПЛА) [9].

Современные системы компьютерного зрения для БПЛА сталкиваются с комплексной задачей

обеспечения надежной детекции и сопровождения объектов в реальном времени при ограниченных вычислительных ресурсах [13–15]. Как показано в работе [16], алгоритм CSRT демонстрирует высокую эффективность в задачах трекинга, однако его работоспособность критически зависит от качества начальной детекции и периодической коррекции положения цели. Это выдвигает особые требования к нейросетевым моделям детекции, которые должны обеспечивать не только высокую точность распознавания, но и эффективную работу в составе гибридной системы.

Анализ современных исследований [1, 2, 17–20] показывает, что существующие реализации YOLO, несмотря на свою популярность, об-

ладают рядом ограничений при использовании в подобных системах. В работах [21, 22] отмечается проблема избыточной вычислительной сложности стандартных архитектур, что делает их неоптимальными для embedded-систем БПЛА. Исследование [23] выявило снижение точности детекции при значительном изменении масштаба объекта, что особенно критично при приближении БПЛА к цели. Авторы статьи [24] подчеркивают трудности интеграции детекторов с алгоритмами трекинга, связанные с неоптимальным форматом выходных данных и отсутствием механизмов адаптации к специфике работы трекера.

Между тем в работах [25, 26] рассмотрен механизм depthwise separable convolutions (DWS) как перспективный метод оптимизации вычислительной нагрузки в сверточных нейросетях. В исследовании [25] доказано, что замена стандартных сверток на DWS-аналоги в детекторах на основе YOLO сокращает параметры модели в 4–11 раз при сохранении 95 % точности. Авторы [26] отмечают, что DWS особенно эффективны для embedded-систем, обеспечивая до 30 % ускорения инференса. В [27] предложена адаптация DWS для задач монообъектной детекции, что актуально для систем с предопределенным целевым классом.

Таким образом, ключевой проблемой нейросетевых моделей детекции в гибридных системах БПЛА является неоптимальное использование вычислительных ресурсов. Как показывают исследования [22, 25], стандартные архитектуры содержат значительную избыточность параметров, обусловленную ориентацией на многоклассовую детекцию, тогда как в рассматриваемом случае требуется распознавание лишь единичного целевого объекта. Это приводит к необоснованному увеличению времени обработки и энергопотребления, что критично для embedded-систем с ограниченными ресурсами. При этом, как отмечается в [26, 27], существующие методы оптимизации, включая DWS, позволяют значительно сократить вычислительную сложность без ощутимой потери точности при работе с предопределенным классом объектов.

Целью авторов настоящей статьи является разработка модифицированной версии YOLOv8m, оптимизированной для работы в составе гибридной системы детекции и трекинга [16] на борту БПЛА с автоматическим наведением. Основные задачи включают сокращение вычислительной сложности модели при сохранении точности детекции, повышение устойчивости к изменению масштаба объекта, обеспечение совместимости с алгоритмом трекинга CSRT, а также адаптацию модели для работы на ресурсоограниченных embedded-платформах.

## Выбор и обоснование базовой архитектуры. Математическая модель и алгоритмическая оптимизация архитектуры

В качестве базовой модели для модификации была выбрана YOLOv8m (medium-версия), что обусловлено оптимальным балансом между точностью детекции ( $mAP@0.5 = 91\%$  на COCO) и вычислительной эффективностью (скорость обработки FPS  $\sim 3,2$  мс на Tesla T4). Данная архитектура демонстрирует лучшие показатели FPS/точность среди семейства YOLOv8 при работе на ресурсоограниченных устройствах, что подтверждено бенчмарками в исследованиях [28–30]. Особое значение имеет модульная структура YOLOv8m, позволяющая производить точечные изменения backbone- и head-частей сети без критического нарушения целостности архитектуры.

В рамках данного исследования проведена корректировка архитектуры YOLOv8m для ее эффективной работы в составе гибридной системы детекции и трекинга на платформе Raspberry Pi 5. Основной акцент сделан на адаптацию сети под задачу детекции единичного объекта, что соответствует требованиям трекера CSRT, работающего только с одним целевым объектом, имеющим наивысший уровень достоверности. Такой подход позволил значительно снизить вычислительную нагрузку на систему без существенного ухудшения точности сопровождения.

Ключевым аспектом модернизации архитектуры YOLOv8m является замена части стандартных сверточных слоев на DWS (рис. 1, а и б). Данный метод, основанный на разделении операции свертки на два последовательных этапа: depthwise (пространственная свертка) и pointwise (объединение каналов), — продемонстрировал значительную эффективность в контексте ресурсоограниченных систем.

Математически процедура DWS выражается как композиция двух операций.

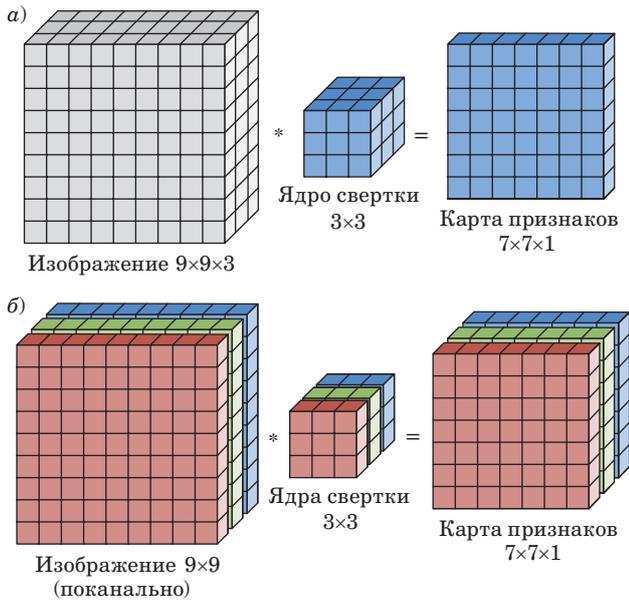
Первая — пространственная свертка, применяемая независимо к каждому каналу входного тензора:

$$DeptConv(W, y)_{(i,j)} = \sum_{k,l}^{K,L} w(k, l) \cdot y(i+k, j+l), \quad (1)$$

где  $w$  — двумерное ядро свертки размерностью  $K \times L$ ;  $y$  — входной тензор высотой и шириной  $H \times W$  и с числом каналов  $C_{in}$ ;  $i, j$  — координаты выходного элемента;  $k, l$  — индексы перемещения по ядру  $w$ .

Вторая — свертка  $1 \times 1$ , комбинирующая каналы:

$$PointConv(W, y)_{(i,j)} = \sum_m^{C_{in}} w_m \cdot y(i, j, m), \quad (2)$$



■ **Рис. 1.** Применения ядра свертки  $3 \times 3$  с использованием классической сверточной обработки (а) и depthwise-обработки (б)  
 ■ **Fig. 1.** Applications of  $3 \times 3$  convolution kernel using classical convolutional processing (a) and depthwise processing (б)

где  $w_m$  — одномерное ядро свертки длиной  $M$  (число входных каналов)  $\times N$  (число выходных каналов);  $i, j$  — координаты пространственного положения на карте признаков;  $m$  — индекс входного канала:  $m \in [1, C_{in}]$ ;  $y(i, j, m)$  — значение входного тензора в точке  $(i, j)$  канала  $m$ .

В стандартных сверточных слоях YOLO ядро фильтра размером  $K \times K$  применяется ко всем  $C_{in}$  входным каналам. Тогда каждое ядро свертки имеет размер  $K \times K \times C_{in}$  (на рис. 1, а —  $9 \times 9 \times 3$ ), а выходной тензор имеет  $(K - 1) \times (K - 2) \times C_{out}$  каналов [31].

В этом случае вычислительная сложность  $O_{conv}$  равна

$$O_{conv} = C_{in} \cdot C_{out} \cdot K^2 \cdot H \cdot W. \quad (3)$$

Таким образом, для входного изображения  $C_{in} = 3$  (RGB-каналы), а в промежуточных слоях, например при входном тензоре размером  $256 \times 56 \times 56$  ( $C_{in} = 256, H = W = 56$ ), применение свертки с  $C_{out} = 512$  фильтров размером  $3 \times 3$  даст выходной тензор  $512 \times 56 \times 56$ . Тогда, согласно формуле (3), для данного случая  $O_{conv} \approx 3,7 \cdot 10^9$ .

В ситуации замены слоев со стандартной сверткой на DWS нет умножения на  $C_{out}$ , поскольку каждый канал обрабатывается независимо [25]. Поэтому формула сложности вычислений (3) примет вид

$$O_{DWS\_S} = C_{in} \cdot K^2 \cdot H \cdot W \approx 7,2 \cdot 10^6. \quad (4)$$

С учетом применения точечной свертки (pointwise convolution), которая выполняет линейное смешивание каналов, каждый из 512 фильтров применяется ко всем 256 каналам в одной точке ( $H, W$ ). В результате поэлементного умножения и последующего суммирования получается один новый канал [32]. При этом коэффициент  $K^2 = 1^2 = 1$ . Вычислительная сложность с учетом этого будет составлять

$$O_{DWS\_P} = C_{in} \cdot C_{out} \cdot H \cdot W \approx 4,1 \cdot 10^8. \quad (5)$$

Для сравнения вычислительной сложности обычной свертки и DWS используем соотношение

$$\frac{O_{conv}}{O_{DWS\_S} + O_{DWS\_P}} = \frac{C_{out} \cdot K_{conv}^2}{K_{DWS\_S}^2 + K_{DWS\_P}^2} \cdot C_{out} \approx 8,8. \quad (6)$$

Из вычислений по формуле (6) видно, что DWS сокращает вычисления в 8,8 раза по сравнению с классической сверткой, разделяя операцию на depthwise-обработку каналов и pointwise-смешивание, что сохраняет точность при значительном ускорении.

Также замечено, что при использовании DWS с ядром  $3 \times 3$  наблюдалось сокращение количества параметров в среднем в 1,73 раза (с 25,9 до 14,9 млн для типового слоя), что подтверждает перспективность данного подхода для embedded-решений.

### Структурные изменения модифицированной YOLOv8m

При модификации YOLOv8m была применена избирательная стратегия замены слоев: преобразованию подверглись преимущественно слои, обрабатывающие признаки среднего и высокого уровней абстракции, где отмечалась наибольшая избыточность вычислительных операций. Важно отметить, что начальные сверточные слои, ответственные за первичную обработку изображения и формирование базовых признаков, были сохранены в исходном виде. Такой подход позволил сохранить информативность низкоуровневых признаков, критически важных для последующих этапов детекции и трекинга.

Выбор конкретных слоев для замены на DWS основывался на результатах комплексного экспериментального исследования. На первом этапе проводился анализ распределения вычислительной нагрузки по слоям модели с использованием

инструментария TorchProfiler. Были идентифицированы слои, вносящие максимальный вклад в общее количество операций (свыше 15 % от 8,5 GFLOPS исходной модели). Затем для этих слоев выполнялось пошаговое ablation-исследование: последовательная замена каждого кандидатного слоя на DWS-аналог с оценкой влияния на mAP@0.5 и FPS на тестовой выборке датасета ONE\_ОБЪЕКТ.

Как показали результаты, замена слоев в блоках C3CA\_3 и C3CA\_9 позволяла достичь 28–35 % сокращения вычислительной сложности при потере точности не более 0,3 процентного пункта для каждого модифицированного слоя. При этом модификация начальных сверточных слоев приводила к существенному падению mAP@0.5 (до 2,5 %), что подтвердило их критическую важность для формирования базовых признаков. Дополнительным критерием отбора служила устойчивость к изменению масштаба объекта — слои, демонстрирующие наименьшее ухудшение точности для объектов малого размера (<50 пикселей), получали приоритет при замене.

Реализация модификаций осуществлялась средствами PyTorch с сохранением совместимости с оригинальной архитектурой YOLOv8 [33]. Для заменяемых слоев применялась специальная процедура инициализации весов: параметры depthwise-части формировались путем декомпозиции предобученных ядер исходной модели,

в то время как pointwise-компоненты инициализировались случайным образом с последующим обучением на целевом датасете ONE\_ОБЪЕКТ (820 изображений целевого объекта, один класс).

С учетом вышеописанного модифицированная архитектура YOLOv8m, где ключевым изменением стала замена стандартных сверточных слоев на DWS в определенных блоках, представлена на рис. 2.

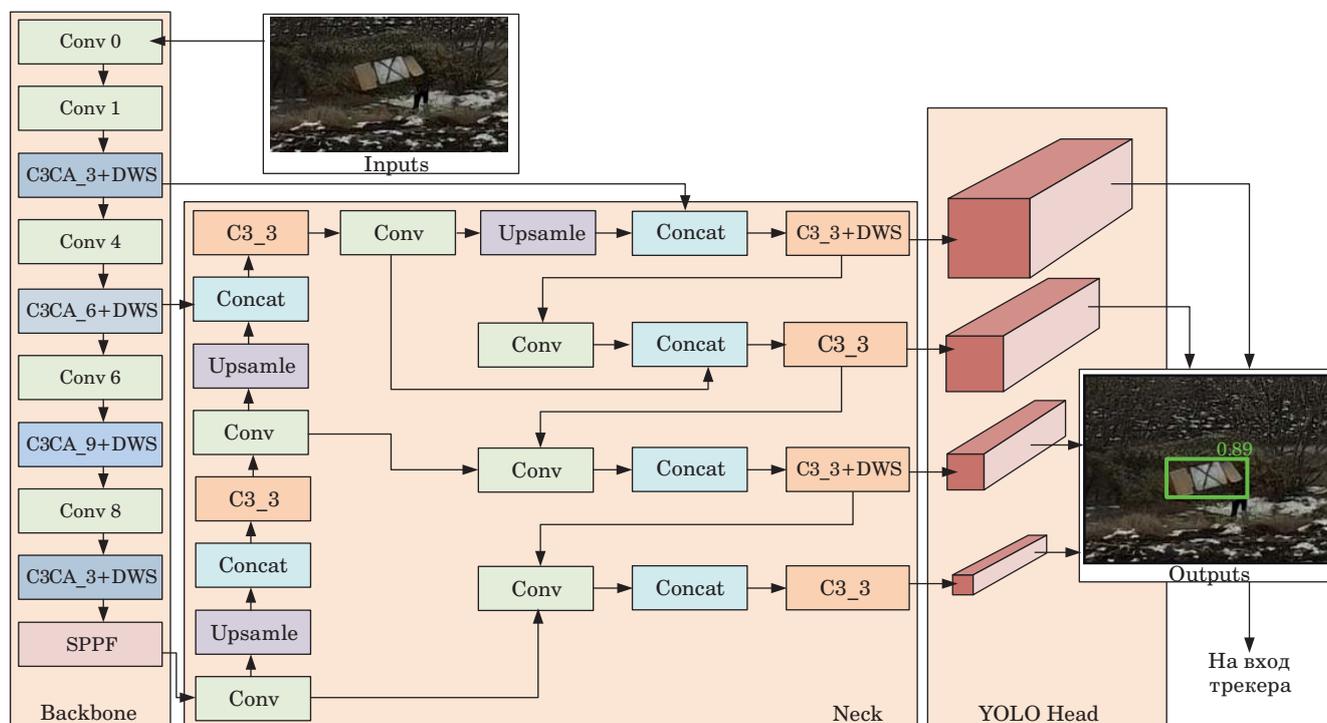
Особенностью предложенной структуры является использование специализированных блоков C3CA [34] в сочетании с DWS-свертками.

Основные преобразования затронули следующие блоки:

- блоки C3CA\_3 подверглись частичной замене (50 %) внутренних сверток на DWS-аналоги;
- блоки C3CA\_6 получили полную замену стандартных сверток на DWS-версии;
- блоки C3CA\_9 получили полную замену стандартных сверток на DWS-версии;
- промежуточные сверточные слои (Conv 1, Conv 4, Conv 6, Conv 8) оптимизированы с сохранением базовой функциональности.

Входной сверточный слой (блок Conv 0) оставлен без изменений для сохранения низкоуровневых признаков. Финальный SSPF-блок остался без изменений для сохранения пространственной пирамидальной структуры.

В ходе оптимизации архитектуры модификации подвергся также блок Neck, отвечающий



■ **Рис. 2.** Структура модифицированной YOLOv8m  
 ■ **Fig. 2.** Modified structure of YOLOv8m

за агрегацию и перераспределение признаков различного масштаба. Основные изменения включают избирательную замену стандартных сверточных слоев на DWS с сохранением критически важных элементов структуры. В верхних слоях Neck, получающих высокоуровневые признаки из Backbone, оставлены обычные свертки для сохранения информативности признаков. Средние и нижние слои, где требования к точности менее строги, были оптимизированы через внедрение DWS-аналогов, что позволило сократить количество параметров на 40 % для каждого модифицированного слоя. Особое внимание уделено блокам C3\_3, где DWS-свертки применены только в residual-ветвях при сохранении основной shortcut-ветви неизменной для стабильности обучения. Первый и последний сверточные слои Neck не подвергались модификации, поскольку они играют ключевую роль в инициализации потока признаков и формировании выходных данных для YOLO Head.

### Результаты исследования

Для валидации модифицированной архитектуры использовался специализированный датасет ONE\_OBJECT, содержащий 820 изображений целевого объекта в разрешении 1920×1080 пикселей. Датасет был создан на основе аэрофотосъемки с БПЛА, выполненной с различных ракурсов и расстояний от 5 до 400 м. В процессе съемки специально варьировалось фоновое окружение для повышения вариативности данных. Особое внимание при формировании датасета уделялось включению объектов малого размера (<50 пикселей), что составило 18 % от общего количества образцов и позволило оценить масштабную устойчивость модели.

Первичные видеоматериалы подвергались комплексной постобработке, включавшей геометрические преобразования с поворотами до ±15° и масштабированием ±20 %, цветокоррек-

цию с вариацией яркости ±30 % и контрастности ±25 %, а также добавление гауссова шума с  $\sigma = 0,05$ . Такой подход к аугментации данных обеспечил необходимое разнообразие обучающей выборки.

Экспериментальные исследования модифицированной версии YOLOv8m проводились с оценкой по ключевым показателям (таблица).

Результаты экспериментального исследования демонстрируют эффективность предложенных архитектурных изменений при сохранении приемлемого уровня точности детекции (рис. 3).

Сравнительный анализ стандартной и модифицированной версий YOLOv8m на датасете ONE\_OBJECT выявил следующие закономерности. По метрике mAP@0.5 наблюдалось незначительное снижение точности на 0,7 процентного пункта (с 94,2 до 93,5 %), что находится в пределах статистической погрешности для задач монообъектной детекции. Наибольшее расхождение в 1,2 % зафиксировано для случаев детекции объектов малого размера (<50 пикселей), что объясняется уменьшением рецептивного поля в оптимизированных слоях.

По производительности модифицированная архитектура показала существенное улучшение ключевых показателей. Скорость обработки возросла на 63,6 % (с 11 до 18 FPS) при тестировании на платформе Raspberry Pi5, что напрямую связано со снижением вычислительной сложности на 32,9 % (с 8,5 до 5,7 GFLOPS).

Анализ ресурсоемкости выявил сокращение количества параметров модели на 37,1 % (с 25,9 до 16,3 млн) и уменьшение требований к памяти на 29,4 % (со 102 до 72 МБ). Эти изменения особенно значимы для систем с ограниченными вычислительными возможностями, где каждый мегабайт памяти имеет критическое значение.

Результаты тестирования также показали снижение среднего энергопотребления системы с 4,8 до 4,4 Вт (−8,3 %) при работе на Raspberry Pi5. Это незначительное улучшение энергоэффективности достигнуто благодаря оптимиза-

- Сравнения моделей YOLOv8m
- Comparison of YOLOv8m models

Параметр	Стандартная YOLOv8m	Модифицированная YOLOv8m	Изменение, %
mAP@0.5	94,2 %	93,5 %	−0,7
FPS (Raspberry Pi5)	11	18	+63,6
GFLOPS	8,5	5,7	−32,9
Количество параметров модели	25,9 млн	16,3 млн	−37,1
Память	102 МБ	72 МБ	−29,4
Средняя энергия потребления	4,8 Вт	4,4 Вт	−8,3



■ **Рис. 3.** Результат сравнительного эксперимента стандартной (слева) и модифицированной (справа) YOLOv8m

■ **Fig. 3.** The result of a comparative experiment of standard (left) and modified (right) YOLOv8m

ции вычислительной нагрузки модели, что особенно важно для автономных систем с ограниченными энергетическими ресурсами, таких как бортовые компьютеры БПЛА.

## Литература

1. **Hanzla M., Jalal A.** Intelligent transportation surveillance via YOLOv9 and NASNet over aerial imagery. *2025 6th International Conference on Advances in Computational Sciences (ICACS)*, Lahore, Pakistan, 2025, pp. 1–8. doi:10.1109/ICACS64902.2025.10937840
2. **Chen J., Wang Q., Cheng H. H., Peng W., Xu W.** A review of vision-based traffic semantic understanding in ITSs. *IEEE Transactions on Intelligent Transportation Systems*, 2022, vol. 23, iss. 11, pp. 19954–19979. doi:10.1109/TITS.2022.3182410
3. **Ganesan M., Chokkalingam B., Kandhasamy S.** Implementation of different road user detection with custom dataset using deep learning algorithm for autonomous vehicle. *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric*

## Заключение

Проведенная модификация YOLOv8m путем замены стандартных сверточных слоев на DWS и оптимизации блоков СЗСА позволила создать эффективную модель для детекции единичного объекта в системах реального времени для БПЛА. Эксперименты показали, что предложенная архитектура сокращает вычислительную нагрузку на 32,9 % (5,7 GFLOPS) и уменьшает объем параметров на 37,1 % (16,3 млн) при сохранении точности детекции (mAP@0.5 = 93,5 %). Наибольшее снижение точности на 1,2 процентного пункта наблюдается для объектов малого размера (<50 пикселей), что открывает перспективы для дальнейшей оптимизации в этом направлении. Также представляет интерес исследование возможностей интеграции с другими алгоритмами трекинга и адаптации архитектуры для задач мультиобъектного сопровождения. Полученные результаты демонстрируют, что избирательное применение DWS-сверток в слоях среднего уровня обеспечивает оптимальный баланс между производительностью и качеством работы модели. Модифицированная версия YOLOv8m особенно перспективна для embedded-систем с ограниченными ресурсами, где критически важны энергоэффективность и скорость обработки, а также открывает новые возможности для создания легковесных гибридных систем детекции и трекинга в различных приложениях компьютерного зрения.

## Финансовая поддержка

Исследование выполнено за счет гранта Российского научного фонда № 25-21-20601, <https://rscf.ru/project/25-21-20601/>.

*Transportation (SEFET)*, Hyderabad, India, 2024, pp. 1–7. doi:10.1109/SEFET61574.2024.10718012

4. **Rajabli N., Flammini F., Nardone R., Vittorini V.** Software verification and validation of safe autonomous cars: A systematic literature review. *IEEE Access*, 2021, vol. 9, pp. 4797–4819. <https://doi.org/10.1109/ACCESS.2020.3048047>
5. **Wang K., Shen C., Li X., Lu J.** Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2025, vol. 26, no. 3, pp. 2880–2896. doi:10.1109/TITS.2025.3532803
6. **Иванов М. Е., Матасов Ю. Ф.** Применение нейросетевого регулятора для повышения качества переходного процесса системы стабилизации высоты полета экраноплана. *Информационно-управляющие системы*, 2024, № 6, с. 27–34. doi:10.31799/1684-8853-2024-6-27-34, EDN: BOBWBN

7. Luu T. A. Application of unmanned aerial vehicle in topographical surveying: A review. *Sustainable Development of Mountain Territories*, 2024, vol. 16, no. 4, pp. 1833–1848. doi:10.21177/1998-4502-2024-16-4-1833-1848
8. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podnebenna S., Belousov E., Volkov V., Ahieiev M., Pohorletskiy D., Zinchenko S., Khudiakov I. Development of a simulation model of a system for automatic regulation of the capacity of a compressor unit. *SAE Technical Papers*, Warrendale, Warrendale, PA, 2022, no. 2022-01-5022, p. 5022. doi:10.4271/2022-01-5022
9. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskiy D., Tkachuk K., Nosov P., Ukrainskiy Ye. Improvement of production processes of the marshalling yard in terms of automatic power control of the compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5063, p. 5063. doi:10.4271/2022-01-5063
10. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskiy D., Tkachuk K., Nosov P., Sumtsov A. Development of a multifactor model for determining energy costs at a compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5079, p. 5079. doi:10.4271/2022-01-5079
11. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podnebenna S., Belousov E., Ahieiev M., Pohorletskiy D., Serhii Z., Khudiakov I. Development of the structure for the technical implementation of the modernized system for automatic regulation of the capacity of the compressor unit at the marshalling yard. *SAE Technical Papers*, 2022, no. 2022-01-5084, p. 5084. doi:10.4271/2022-01-5084
12. Volodarets M., Gritsuk I., Ukrainskiy Y., Shein V., Stepanov O., Khudiakov I., Ahieiev M., Vychuzhannin V., Smyrnov O., Saraiev O. Development of the analytical system for vehicle operating conditions management in the V2I information complex using simulation modeling. *Eastern-European Journal of Enterprise Technologies*, 2020, no. 3 (107), pp. 6–16. doi:10.15587/1729-4061.2020.215006
13. Redmon J., Farhadi A. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. doi:10.48550/arXiv.1804.02767
14. Bochkovskiy A., Wang C.-Y., Liao H.-Y. M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. doi:10.48550/arXiv.2004.10934
15. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. doi:10.48550/arXiv.2207.02696
16. Сацюк А. В., Володарец Н. В. Оптимизация параметров трекера CSRT в системе слежения беспилотных летательных аппаратов. *Устойчивое развитие горных территорий*, 2025, т. 17, № 2. <https://naukagor.ru/ru-ru/about/textsofarticles> (дата обращения: 07 июля 2025). (В печати.)
17. Кориков А. М., Тран В. Т. Нейросетевой анализ видеопотока, поступающего с камеры беспилотного летательного аппарата, для компенсации ошибок системы навигации и управления БПЛА. *Телекоммуникации*, 2023, № 7, с. 13–23. doi:10.31044/1684-2588-2023-0-7-13-23
18. Viswanatha V., Chandana R. K., Ramachandra A. C. Real time object detection system with YOLO and CNN models: A review. *arXiv preprint arXiv:2208.773*, 2022. doi:10.48550/arXiv.2208.00773
19. Yuan K., Zhuang X., Schaefer G., Feng J., Guan L., Fang H. Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2021, vol. 14, pp. 7422–7434. doi:10.1109/JSTARS.2021.3098678
20. Gallagher J. E., Oughton E. J. Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications, and challenges. *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), IEEE Access*, 2025, vol. 13, pp. 7366–7395. doi:10.1109/ACCESS.2025.3526458
21. Xu S., Xue Y. Lightweight object detection for UAVs using compressed feature fusion and attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, pp. 1–14. doi:10.1109/TGRS.2023.3267421
22. Satsuk A. V., Volodarets N. V., Krupskiy D. I. Acceleration of neural network for UAV operation in real time. *Вестник Донецкой академии транспорта*, 2025, № 1, с. 6–14.
23. Laghari A. A., Jumani A. K., Laghari R. A., Li H., Karim S., Khan A. A. Unmanned aerial vehicles advances in object detection and communication security review. *Cognitive Robotics*, 2024, vol. 4, pp. 73–82. doi:10.1016/j.cogr.2024.07.002
24. Wang Q., Gao J., Xing J., Zhang M. DETR4D: Transformer-based multi-object tracking for autonomous driving with depth-aware detection. *IEEE Robotics and Automation Letters*, 2023, vol. 8, no. 4, pp. 1931–1938. doi:10.1109/LRA.2023.3246839
25. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. doi:10.48550/arXiv.1704.04861
26. Hasan M. A., Dey K. Depthwise separable convolutions with deep residual convolutions. *arXiv preprint arXiv:2411.07544*, 2024. doi:10.48550/arXiv.2411.07544
27. Zhao X., Wang L., Zhang Y., Han X., Deveci M., Parmar M. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 2024, vol. 57, art. no. 99, p. 43. doi:10.1007/s10462-024-10721-6
28. Wang J.-W., Cao Y., Guo Z.-K., Xu C. Research on long-distance snow depth measurement method based on improved YOLOv8. *IEEE Access*, 2025, vol. 13, pp. 55370–55380. doi:10.1109/ACCESS.2025.3551727

29. Varghese R., Sambath M. Ultralytics Yolov8: A novel object detection algorithm with enhanced performance and robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India, 2024, pp. 1–6. doi:10.1109/ADICS58448.2024.10533619
30. Alqahtani D. K., Cheema M. A., Toosi A. N. Benchmarking deep learning models for object detection on edge computing devices. *International Conference on Service-Oriented Computing*, LNCS, Singapore, Springer Nature Singapore, 2025, vol. 15404, pp. 142–150. doi:10.1007/978-981-96-0805-8\_11
31. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, pp. 436–444. doi:10.1038/nature14539
32. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
33. Wang X., Li Z., Yang J. PyTorch-YOLO-Nano. A framework for efficient deployment of modified YOLO architectures on embedded GPUs. *Journal of Real-Time Image Processing*, 2025, no. 22(1), pp. 45–60. doi:10.1007/s11554-024-01458-0
34. Chen L., Wang Z., Liu F. C3Fusion: Lightweight cross-channel and spatial attention for real-time object detection on edge devices. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023, no. 46(5), pp. 1123–1136. doi:10.1109/TPAMI.2024.3385210

UDC 004.8+004.93+629.7

doi:10.31799/1684-8853-2025-4-36-44

EDN: YKQVJU

### Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system

A. V. Satsiuk<sup>a</sup>, PhD, Tech., Associate Professor, orcid.org/0009-0006-7228-8279

N. V. Volodarets<sup>b</sup>, PhD, Tech., Associate Professor, orcid.org/0000-0002-8526-4800, volodarets.nikita@yandex.ru

<sup>a</sup>Donetsk Institute of Railway Transport, 6, Gornaya St., 283018, Donetsk, Russian Federation

<sup>b</sup>Donetsk Academy of Transport, 7, Dzerzhinsky Ave., 283086, Donetsk, Russian Federation

**Introduction:** Modern computer vision systems for UAVs face the problem of reliable detection and tracking in real time with limited resources of embedded platforms, especially when integrating neural network detectors with tracking algorithms. Existing YOLO implementations, despite their popularity, have drawbacks: excessive computational complexity due to the focus on multi-class detection and non-optimal interaction with tracking algorithms. **Purpose:** To develop an optimized version of YOLOv8 for a hybrid detection and an on-board tracking system for a UAV with automatic guidance, aimed at reducing computational complexity while maintaining accuracy and adapting to resource-limited platforms. **Results:** The study is based on experiments with a modified YOLOv8m, evaluated on an embedded platform (Raspberry Pi 5) and a specialized ONE\_OBJECT dataset. We develop a modified version of YOLOv8m with selective replacement of standard convolutional layers with depthwise separable convolutions in C3CA blocks and some Neck layers. As a result of experimental studies, we achieve a reduction in computational complexity by 32.9% (from 8.5 to 5.7 GFLOPS), the number of parameters by 37.1% (from 25.9 million to 16.3 million) and memory requirements by 29.4% (from 102 to 72 MB). The processing speed on Raspberry Pi 5 increases by 63.6% (from 11 to 18 FPS) while maintaining high detection accuracy mAP@0.5 at 93.5% (a drop of only 0.7 percentage points relative to the base model when tested on the ONE\_OBJECT dataset). The greatest decrease in accuracy (1.2 percentage points) is observed for small objects (less than 50 pixels). **Practical relevance:** The developed YOLOv8m modification has been successfully integrated with the CSRT tracking algorithm. Consequently, it becomes possible to create efficient hybrid automatic guidance systems for UAVs. The proposed solution is especially promising for embedded systems with limited resources. In addition, the obtained results open up new possibilities for creating energy-efficient real-time computer vision systems.

**Keywords** – YOLOv8, depthwise separable convolutions, UAV, hybrid detection, CSRT tracking, modification, neural network optimization.

**For citation:** Satsiuk A. V., Volodarets N. V. Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 36–44 (In Russian). doi:10.31799/1684-8853-2025-4-36-44, EDN: YKQVJU

#### Financial support

The study was supported by the Russian Science Foundation, grant No. 25-21-20601, <https://rscf.ru/project/25-21-20601/>.

#### Reference

- Hanzla M., Jalal A. Intelligent transportation surveillance via YOLOv9 and NASNet over aerial imagery. *2025 6th International Conference on Advances in Computational Sciences (ICACS)*, Lahore, Pakistan, 2025, pp. 1–8. doi:10.1109/ICACS64902.2025.10937840
- Chen J., Wang Q., Cheng H. H., Peng W., Xu W. A review of vision-based traffic semantic understanding in ITSs. *IEEE Transactions on Intelligent Transportation Systems*, 2022, vol. 23, iss. 11, pp. 19954–19979. doi:10.1109/TITS.2022.3182410
- Ganesan M., Chokkalingam B., Kandhasamy S. Implementation of different road user detection with custom dataset using deep learning algorithm for autonomous vehicle. *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, Hyderabad, India, 2024, pp. 1–7. doi:10.1109/SEFET61574.2024.10718012
- Rajabli N., Flammini F., Nardone R., Vittorini V. Software verification and validation of safe autonomous cars: A systematic literature review. *IEEE Access*, 2021, vol. 9, pp. 4797–4819. <https://doi.org/10.1109/ACCESS.2020.3048047>

5. Wang K., Shen C., Li X., Lu J. Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2025, vol. 26, no. 3, pp. 2880–2896. doi:10.1109/TITS.2025.3532803
6. Ivanov M. E., Matasov Yu. F. Neuro-controller for improving the quality of the transient process of the WIG-craft altitude stabilization system. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2024, no. 6, pp. 27–34 (In Russian). doi:10.31799/1684-8853-2024-6-27-34, EDN: BOBWBN
7. Luu T. A. Application of unmanned aerial vehicle in topographical surveying: A review. *Sustainable Development of Mountain Territories*, 2024, vol. 16, no. 4, pp. 1833–1848. doi:10.21177/1998-4502-2024-16-4-1833-1848
8. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podnebenna S., Belousov E., Volkov V., Ahieiev M., Pohorletskiy D., Zinchenko S., Khudiakov I. Development of a simulation model of a system for automatic regulation of the capacity of a compressor unit. *SAE Technical Papers*, Warrendale, Warrendale, PA, 2022, no. 2022-01-5022, p. 5022. doi:10.4271/2022-01-5022
9. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskiy D., Tkachuk K., Nosov P., Ukrainskiy Ye. Improvement of production processes of the marshalling yard in terms of automatic power control of the compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5063, p. 5063. doi:10.4271/2022-01-5063
10. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskiy D., Tkachuk K., Nosov P., Sumtsov A. Development of a multifactor model for determining energy costs at a compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5079, p. 5079. doi:10.4271/2022-01-5079
11. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podnebenna S., Belousov E., Ahieiev M., Pohorletskiy D., Serhii Z., Khudiakov I. Development of the structure for the technical implementation of the modernized system for automatic regulation of the capacity of the compressor unit at the marshalling yard. *SAE Technical Papers*, 2022, no. 2022-01-5084, p. 5084. doi:10.4271/2022-01-5084
12. Volodarets M., Gritsuk I., Ukrainskiy Y., Shein V., Stepanov O., Khudiakov I., Ahieiev M., Vychuzhanin V., Smyrnov O., Saraiev O. Development of the analytical system for vehicle operating conditions management in the V2I information complex using simulation modeling. *Eastern-European Journal of Enterprise Technologies*, 2020, no. 3 (107), pp. 6–16. doi:10.15587/1729-4061.2020.215006
13. Redmon J., Farhadi A. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. doi:10.48550/arXiv.1804.02767
14. Bochkovskiy A., Wang C.-Y., Liao H.-Y. M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. doi:10.48550/arXiv.2004.10934
15. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. doi:10.48550/arXiv.2207.02696
16. Satsiuk A. V., Volodarets N. V. Optimization of CSRT tracker parameters in unmanned aerial vehicle tracking system. *Sustainable Development of Mountain Territories*, 2025, vol. 17, no. 2. Available at: <https://naukagor.ru/ru-ru/about/textsofarticles> (accessed 07 July 2025) (In Russian). (In print.)
17. Korikov A. M., Tran V. T. Neural network analysis of the video stream coming from the camera of an unmanned aerial vehicle to compensate for errors in the UAV navigation and control system. *Telecommunications*, 2023, no. 7, pp. 13–23. doi:10.31044/1684-2588-2023-0-7-13-23
18. Viswanatha V., Chandana R. K., Ramachandra A. C. Real time object detection system with YOLO and CNN models: A review. *arXiv preprint arXiv:2208.773*, 2022. doi:10.48550/arXiv.2208.00773
19. Yuan K., Zhuang X., Schaefer G., Feng J., Guan L., Fang H. Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2021, vol. 14, pp. 7422–7434. doi:10.1109/JSTARS.2021.3098678
20. Gallagher J. E., Oughton E. J. Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications, and challenges. *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, *IEEE Access*, 2025, vol. 13, pp. 7366–7395. doi:10.1109/ACCESS.2025.3526458
21. Xu S., Xue Y. Lightweight object detection for UAVs using compressed feature fusion and attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, pp. 1–14. doi:10.1109/TGRS.2023.3267421
22. Satsuk A. V., Volodarets N. V., Krupskiy D. I. Acceleration of neural network for UAV operation in real time. *Vestnik Doneckoj Akademii Transporta*, 2025, no. 1, pp. 6–14.
23. Laghari A. A., Jumani A. K., Laghari R. A., Li H., Karim S., Khan A. A. Unmanned aerial vehicles advances in object detection and communication security review. *Cognitive Robotics*, 2024, vol. 4, pp. 73–82. doi:10.1016/j.cogr.2024.07.002
24. Wang Q., Gao J., Xing J., Zhang M. DETR4D: Transformer-based multi-object tracking for autonomous driving with depth-aware detection. *IEEE Robotics and Automation Letters*, 2023, vol. 8, no. 4, pp. 1931–1938. doi:10.1109/LRA.2023.3246839
25. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. doi:10.48550/arXiv.1704.04861
26. Hasan M. A., Dey K. Depthwise separable convolutions with deep residual convolutions. *arXiv preprint arXiv:2411.07544*, 2024. doi:10.48550/arXiv.2411.07544
27. Zhao X., Wang L., Zhang Y., Han X., Deveci M., Parmar M. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 2024, vol. 57, art. no. 99, p. 43. doi:10.1007/s10462-024-10721-6
28. Wang J.-W., Cao Y., Guo Z.-K., Xu C. Research on long-distance snow depth measurement method based on improved YOLOv8. *IEEE Access*, 2025, vol. 13, pp. 55370–55380. doi:10.1109/ACCESS.2025.3551727
29. Varghese R., Sambath M. Ultralytics Yolov8: A novel object detection algorithm with enhanced performance and robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India, 2024, pp. 1–6. doi:10.1109/ADICS58448.2024.10533619
30. Alqahtani D. K., Cheema M. A., Toosi A. N. Benchmarking deep learning models for object detection on edge computing devices. *International Conference on Service-Oriented Computing*, LNCS, Singapore, Springer Nature Singapore, 2025, vol. 15404, pp. 142–150. doi:10.1007/978-981-96-0805-8\_11
31. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, pp. 436–444. doi:10.1038/nature14539
32. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
33. Wang X., Li Z., Yang J. PyTorch-YOLO-Nano. A framework for efficient deployment of modified YOLO architectures on embedded GPUs. *Journal of Real-Time Image Processing*, 2025, no. 22(1), pp. 45–60. doi:10.1007/s11554-024-01458-0
34. Chen L., Wang Z., Liu F. C3Fusion: Lightweight cross-channel and spatial attention for real-time object detection on edge devices. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023, no. 46(5), pp. 1123–1136. doi:10.1109/TPAMI.2024.3385210



## Многокритериальный анализ методов оптимизации веб-страниц и их влияние на ранжирование в поисковых системах

С. В. Жуков<sup>а</sup>, аспирант, [orcid.org/0009-0003-8954-8565](https://orcid.org/0009-0003-8954-8565)

О. А. Ковалева<sup>а</sup>, доктор техн. наук, доцент, [orcid.org/0000-0003-0735-6205](https://orcid.org/0000-0003-0735-6205), [solomina-oa@yandex.ru](mailto:solomina-oa@yandex.ru)

С. В. Ковалев<sup>а</sup>, доктор техн. наук, доцент, [orcid.org/0000-0002-5961-7561](https://orcid.org/0000-0002-5961-7561)

<sup>а</sup>Тамбовский государственный университет им. Г. Р. Державина, Интернациональная ул., 33, Тамбов, 392000, РФ

**Введение:** скорость загрузки веб-страниц является одним из значимых факторов, влияющих на ранжирование сайтов в поисковых системах. Медленная загрузка веб-страницы негативно сказывается на пользовательском опыте, следовательно, может приводить к потере потенциальных клиентов и снижению конверсий. В условиях растущей конкуренции оптимизация скорости становится необходимым элементом разработки эффективных веб-ресурсов, напрямую влияющим на успех сайта в сети. **Цель:** провести многокритериальный анализ и определить наиболее эффективные методы оптимизации скорости загрузки веб-страницы для определенного сайта. **Результаты:** в ходе исследования разработана иерархическая модель критериев для оценки эффективности различных методов оптимизации, включающая серверные, сетевые и клиентские критерии скорости загрузки веб-страницы. Проведен анализ методов оптимизации веб-страниц на примере конкретного веб-сайта. Его результат показал, что наибольший эффект достигается за счет применения метода отложенной загрузки медиаконтента («ленивой загрузки»). Также было выявлено, что некоторые методы оптимизации не имеют значимого влияния на серверные и сетевые критерии производительности сайта. На основании полученных данных создан второй сайт с применением наиболее эффективных методов оптимизации веб-страницы. В результате сравнительного эксперимента оптимизированный сайт показал значительное улучшение по всем ключевым показателям, таким как количество визитов, продолжительность нахождения на сайте, глубина просмотра страниц, снижение показателя отказов. **Практическая значимость:** полученные результаты и предложенная модель позволяют разработчикам принимать обоснованные решения по выбору наиболее эффективных методов оптимизации скорости загрузки веб-страниц.

**Ключевые слова** – скорость загрузки веб-страниц, оптимизация веб-страниц, многокритериальный анализ, метод парных сравнений, поисковая оптимизация, поведенческие факторы, веб-разработка.

**Для цитирования:** Жуков С. В., Ковалева О. А., Ковалев С. В. Многокритериальный анализ методов оптимизации веб-страниц и их влияние на ранжирование в поисковых системах. *Информационно-управляющие системы*, 2025, № 4, с. 45–57. doi:10.31799/1684-8853-2025-4-45-57, EDN: LPNYCL

**For citation:** Zhukov S. V., Kovaleva O. A., Kovalev S. V. Multi-criteria analysis of web page optimization methods and their impact on search engine rankings. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 45–57 (In Russian). doi:10.31799/1684-8853-2025-4-45-57, EDN: LPNYCL

### Введение

Интернет претерпел масштабные изменения за годы своего развития. В статье [1] авторы отмечают, что за это время существенно изменились подходы к организации и представлению информации, разработаны новые технологии и принципы, такие как Web3. Современный Интернет стал не просто средой передачи данных, а многофункциональной платформой, которая оказывает значительное влияние на экономику и рекламную индустрию.

Одной из важнейших задач, стоящих перед разработчиками, остается повышение позиций сайта в результатах поисковой выдачи. В работе [2] отмечается, что грамотная SEO (search engine optimization) способствует не только росту органического трафика, но и улучшению пользова-

тельского опыта и повышению доверия к сайту. Автор выделяет важность таких факторов, как релевантность контента, качество обратных ссылок, скорость загрузки страниц, адаптивность сайта для мобильных устройств. Эти меры позволяют сайту соответствовать требованиям поисковых систем и обеспечивать высокий уровень удовлетворенности пользователей.

Одним из подходов к повышению позиций сайта является оптимизация контента и html-кода сайта, например использование микроразметок. Авторы статьи [3] отмечают, что такие разметки позволяют включать в контент страницы данные внешних ресурсов. Но не менее важную роль в SEO играют и поведенческие факторы, основанные на предпочтениях пользователей. Как показало исследование [4], пользователи, оценивая веб-ресурсы, обращают внимание не только

на контент, такой как отзывы или информация о специалистах, но и на технические характеристики, особенно на скорость загрузки страниц.

Таким образом, обеспечение высокой скорости загрузки становится одним из ключевых аспектов разработки успешного веб-сайта. В условиях растущей конкуренции и повышающихся требований к пользовательскому опыту медленная загрузка веб-сайта представляет серьезную проблему, влияющую на удержание пользователей, конверсию и эффективность взаимодействия с ресурсом. Это подтверждают авторы статьи [5], считая скорость загрузки веб-страниц одним из факторов, влияющих на поисковую оптимизацию.

Целью исследования является проведение многокритериального анализа и выбор оптимальных методов оптимизации скорости загрузки веб-страницы.

Для достижения цели исследования необходимо решить следующие задачи:

- сформулировать методологический подход к оценке влияния методов оптимизации на скорость загрузки веб-страниц;

- разработать иерархическую модель критериев, которая позволит структурировать и приоритизировать факторы, влияющие на скорость загрузки веб-страницы;

- провести анализ методов, влияющих на скорость загрузки веб-страницы, и оценить их эффективность с помощью составленной модели на примере конкретного сайта;

- подтвердить эффективность методов, предложенных иерархической моделью критериев, путем сравнения данных посещаемости двух сайтов – оптимизированного и не оптимизированного.

### Общая постановка задачи

В рамках данной работы рассматривается методологический подход к оценке влияния методов оптимизации на скорость загрузки страниц веб-сайта. Основная цель модели – определить, как конкретный метод оптимизации повлиял на скорость загрузки страницы и насколько сильно изменились показатели производительности.

Для формального описания процесса оценки методов оптимизации веб-страниц определим ключевые компоненты, которые будут использоваться при построении модели анализа эффективности внесенных изменений.

$Q = \{q_a, a = \overline{1, b}\}$  – множество веб-сайтов, анализируемых в рамках исследования.

$U = \{u_o, o = \overline{1, d}\}$  – множество страниц сайтов, для которых оценивается влияние внесенных изменений.

$X = \{x_i, i = \overline{1, n}\}$  – множество альтернатив – методов оптимизации, которые могут быть применены для улучшения скорости загрузки.

$K = \{K_j, j = \overline{1, m}\}$  – множество количественных и качественных критериев, по которым оцениваются альтернативы.

$K_j = \{k_{jt}, t = \overline{1, s_j}, j = \overline{1, m}\}$  – множество частных критериев (подкритериев), характеризующих каждый из критериев.

$Y$  – диапазон значений для каждого критерия, включая такие параметры, как время загрузки в миллисекундах, количество запросов и размер страницы.

$W = \{w_j, j = \overline{1, m}\}$  – коэффициенты относительной важности критериев.

$w_j = \{w_{jt}, t = \overline{1, s_j}, j = \overline{1, m}\}$  – коэффициенты относительной важности частных критериев.

$C \rightarrow [-100, 100]$  – набор функций и инструментов, позволяющих перевести показатели критериев в нормированную шкалу от –100 до 100.

Таким образом, процесс оценки влияния методов оптимизации на скорость загрузки веб-страницы можно описать в виде модели

$$M_{\text{опт}} = (Q, U, X, Y, W, K, C),$$

где результатом работы модели является измерение изменения скорости загрузки страницы после применения конкретного метода оптимизации.

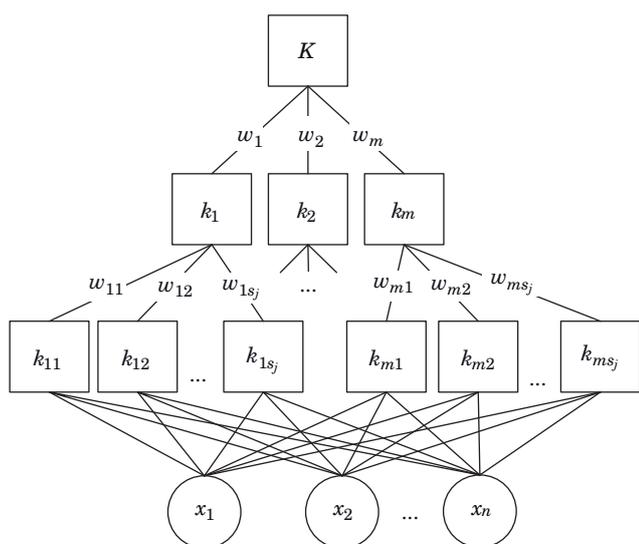
Описание процесса анализа включает несколько последовательных этапов. Сначала производится выбор исследуемого веб-сайта ( $Q$ ) и конкретной страницы ( $U$ ). Измеряем количественные и качественные критерии ( $K$ ) до применения метода оптимизации и после ( $X$ ). Полученные данные ( $Y$ ) подвергаются обработке с применением специальных инструментов ( $C$ ), позволяющих привести их к единой шкале оценки. Далее выполняется учет весов критериев ( $W$ ), что необходимо для получения объективных результатов. Итоговый анализ позволяет определить, насколько каждый метод оптимизации повлиял на скорость загрузки страницы, выявить возможные улучшения или ухудшения, и оценить степень изменений. Иерархия критериев и подкритериев представлена на рис. 1.

Пусть  $f(x)$  – целевая функция, которая обеспечивает выбор наилучших альтернатив:

$$f(x) = \max(f(x_1), f(x_2), \dots, f(x_n)),$$

$$f(x) \rightarrow [-100, 100],$$

где  $f(x_i)$  – результирующий вектор оценки альтернативы  $x_i \in X$  ( $x_i$  – конкретный метод опти-



■ **Рис. 1.** Иерархическая модель критериев  
 ■ **Fig. 1.** Hierarchical criteria model

мизации веб-страницы) в соответствии с интегральным критерием  $K$ , т. е.  $f(x_i) \rightarrow K(x_i)$ .

Таким образом, целью модели является определение метода оптимизации  $x_i$ , который максимизирует положительное влияние на скорость загрузки, при этом минимизируя возможное негативное воздействие. Влияние каждого метода оптимизации оценивается на основе нескольких критериев, определяющих производительность на различных уровнях:

$$K(x_i) = (p(x_i), W),$$

где  $K(x_i)$  – интегральная оценка альтернатив (метода оптимизации)  $x_i$ ;  $p(x_i)$  – разница между значением метрики до и после применения метода оптимизации;  $W$  – веса критериев, определяющие их относительную значимость. Таким образом, итоговая оценка  $K(x_i)$  зависит от измеренных данных о скорости загрузки и приоритетности различных критериев.

Ограничения интегральной оценки гарантируют, что модель корректно интерпретирует случаи, когда метод оптимизации негативно влияет на скорость загрузки:

$$f(x_i) \geq -100 \text{ при условии } p(x_i) \geq -100.$$

Так как критерии имеют различное значение для общего результата, каждому из них присваивается вес:

$$w_j \geq 0, j = \overline{1, m}, \sum_{j=1}^m w_j = 1,$$

$$w_{jt} \geq 0, t = \overline{1, s_j}, \sum_{t=1}^{s_j} w_{jt} = 1,$$

где  $w_j$  – вес критерия  $j$ , отражающий его значимость в общей оценке метода оптимизации;  $m$  – общее количество критериев;  $s_j$  – количество подкритериев для критерия  $j$ ;  $w_{jt}$  – вес подкритерия  $t$  внутри критерия  $j$ .

Эти условия обеспечивают то, что каждый вес положителен или равен нулю ( $w_j \geq 0$ ) и сумма всех весов критериев равна 1, т. е. все критерии в совокупности дают 100 % значимости.

Мы выбираем такие методы оптимизации  $x_i$ , которые максимизируют интегральную оценку, т. е. наиболее положительно влияют на скорость загрузки:

$$X^* = \{x_i | K(x_i) \geq F, i = \overline{1, n}\},$$

где  $X^*$  – множество оптимальных методов оптимизации, отобранных на основе их влияния;  $K(x_i)$  – интегральная оценка метода оптимизации;  $F$  – пороговое значение, определяющее те методы оптимизации, которые считаются положительно влияющими на скорость загрузки.

Таким образом, отбираются все методы оптимизации, у которых  $K(x_i)$  превышает заранее заданный порог  $F$ , выбранный пользователем. Например, если  $F = 10$ , то будут выбраны все методы оптимизации, улучшающие скорость загрузки более чем на 10 единиц.

Предложенная модель представляет собой частный случай многокритериального анализа (МКА), используемого для принятия решений в условиях нескольких оцениваемых параметров. В отличие от классического подхода, в котором выбирается единственная оптимальная альтернатива, данный метод позволяет отбирать сразу несколько методов оптимизации, обладающих наибольшим положительным влиянием на скорость загрузки веб-страниц. Это расширяет применимость модели и делает ее полезной в случаях, когда оптимизация требует комбинированного внесения изменений.

### Применение разработанной иерархической модели критериев для выбора методов оптимизации скорости загрузки веб-страниц

Многокритериальный анализ представляет собой методологию принятия решений, ориентированную на учет нескольких критериев, которые могут быть конфликтующими. Этот подход используется для выбора оптимального решения

из множества альтернатив. В настоящем исследовании для реализации МКА применяется метод анализа иерархий Т. Саати, позволяющий структурировать задачу в виде иерархии критериев и альтернатив. В контексте повышения скорости загрузки веб-страниц МКА позволяет оценить эффективность различных методов оптимизации. Например, такие методы, как сжатие изображений или минимизация кода JavaScript и CSS-кода, могут различаться по уровню их влияния на скорость формирования страницы на веб-сервере и скорость отрисовки страницы в браузере клиента.

Воспользуемся критериями и иерархией, предложенной в работе [6]. Модель, описанная в данной работе, позволяет оценивать влияние любых изменений на веб-ресурсе на общую скорость загрузки, включая те, которые на первый взгляд с ней не связаны, — например, внедрение защиты от SQL-инъекций, как показано в работе [7]. На первом уровне иерархической модели находится ключевой критерий — скорость загрузки страницы ( $K$ ). Она определяется тремя сущностями, которые станут критериями второго уровня: серверным уровнем ( $k_1$ ), сетевым уровнем ( $k_2$ ), клиентским уровнем ( $k_3$ ).

Серверный уровень  $k_1$  отвечает за формирование страницы на сервере, включая выполнение скриптов и работу с базой данных. Сетевой уровень  $k_2$  описывает этапы передачи данных между клиентом и сервером такие, как установление соединения, передача запроса и получение ответа. Клиентский уровень  $k_3$  охватывает обработку страницы в браузере, включая отображение контента и взаимодействие пользователя с интерфейсом.

Серверный уровень включает такие метрики, как время генерации страницы, использование кеша, размер сжатой страницы и количество ошибок. Также учитываются количество запросов к базе данных и время их выполнения. Эти метрики не ограничиваются только реляционными базами данных и при необходимости могут быть применены к нереляционным базам данных. Эти показатели помогают оценить эффективность формирования ответа на сервере и выявить возможные узкие места. Количество ошибок и размер кеша не используются в иерархической модели критериев, поскольку они не оказывают прямого влияния на скорость загрузки страницы, а лишь помогают выявлять проблемы при ее формировании.

На сетевом уровне ключевыми метриками являются время поиска доменного имени, время установки защищенного соединения, задержка при установке соединения и скорость передачи данных. Эти показатели характеризуют сетевые взаимодействия между сервером и клиентом, ко-

торые напрямую влияют на общее время загрузки страницы.

Клиентский уровень включает метрики, оценивающие пользовательский опыт, такие, как FCP (First Contentful Paint — время первой отрисовки веб-страницы), LCP (Largest Contentful Paint — отрисовка крупнейшего элемента), TBT (Total Blocking Time — общий блокирующий интервал), CLS (Cumulative Layout Shift — кумулятивное смещение макета страницы) и SI (Speed Index — скорость отображения видимого контента). Эти параметры показывают, насколько быстро и плавно пользователь может взаимодействовать с загружаемым сайтом.

Иерархическая модель критериев представлена в виде схемы на рис. 2, где наглядно показаны уровни и метрики, обеспечивающие комплексную оценку всех этапов загрузки веб-страницы.

Итоговый балл изменения скорости будет рассчитан как взвешенная сумма критериев всех уровней. Представим модель в виде структуры формул:

$$K = w_1 \cdot k_1 + w_2 \cdot k_2 + w_3 \cdot k_3,$$

$$k_1 = w_{11} \cdot k_{11} + w_{12} \cdot k_{12} + w_{13} \cdot k_{13} + w_{14} \cdot k_{14},$$

$$k_2 = w_{21} \cdot k_{21} + w_{22} \cdot k_{22} + w_{23} \cdot k_{23} + w_{24} \cdot k_{24},$$

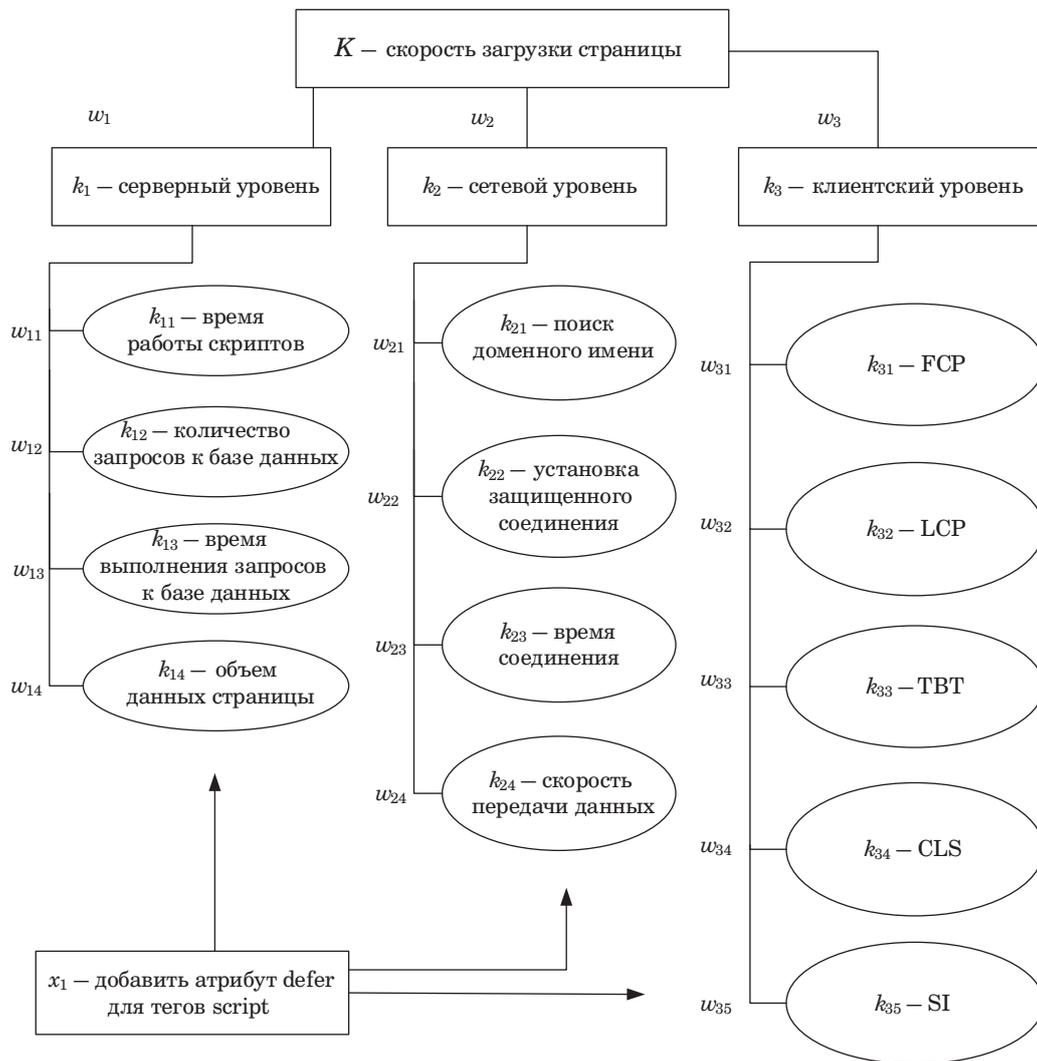
$$k_3 = w_{31} \cdot k_{31} + w_{32} \cdot k_{32} + w_{33} \cdot k_{33} + w_{34} \cdot k_{34} + w_{35} \cdot k_{35}.$$

Следующим этапом построения иерархической модели критериев является распределение весов. Для частных критериев, относящихся к клиентскому уровню  $k_3$ , будут применены веса, предложенные в десятой версии инструмента Lighthouse, поскольку модель опирается на его метрики:

$$w_{31} = 0,1, w_{32} = 0,25, w_{33} = 0,3, w_{34} = 0,25, w_{35} = 0,1.$$

Для расчета весов частных критериев серверного и сетевого уровней применим метод парных сравнений. Он используется для оценки относительной важности или предпочтительности элементов. Пусть есть  $n$  элементов, которые нужно сравнить. Мы строим квадратную матрицу парных сравнений  $A$ , где каждая ячейка  $a_{jt}$  показывает, насколько элемент  $j$  важнее элемента  $t$ :

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ \frac{1}{a_{12}} & 1 & a_{23} & \dots & a_{2n} \\ a_{13} & \frac{1}{a_{23}} & 1 & \dots & a_{3n} \\ & & & \ddots & \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \frac{1}{a_{3n}} & \dots & 1 \end{bmatrix}.$$



■ **Рис. 2.** Иерархическая модель критериев  
 ■ **Fig. 2.** Hierarchical criteria model

Будем использовать шкалу от 1 до 9 для значений  $a_{jt}$ , где 1 – одинаковая значимость  $j$  и  $t$ ; 3 – некоторое преобладание значимости  $j$  над  $t$ ; 5 – существенная или сильная значимость  $j$  против  $t$ ; 7 – очень сильная или очевидная значимость  $j$  над  $t$ ; 9 – абсолютная значимость  $j$  против  $t$ ; 2, 4, 6, 8 – промежуточные значения между соседними значениями шкалы. Обратные значения в матрице парных сравнений рассчитываются как величины, обратные соответствующим прямым значениям ( $1/a_{jt}$ ). После составления матрицы ее необходимо проверить на согласованность. Подробнее о согласованности матриц изложено в работах [8, 9]. Так как оценка проводится на основании одной матрицы парных сравнений, составленной авторами статьи, для нахождения весов каждого элемента используется метод, основанный на корне четвертой степени от произведений элементов строки, нормированных на общую сумму.

Применим описанный метод для оценки относительной значимости частных критериев. Так, частный критерий серверного уровня  $k_{21}$  оказался более значимым, чем  $k_{22}$  и  $k_{23}$ , поскольку без корректного выполнения процесса поиска доменного имени дальнейшие шаги невозможны. Если сравнивать  $k_{21}$  с  $k_{24}$ , становится очевидным, что  $k_{24}$  более важен, так как основная масса времени уходит на загрузку данных, тогда как остальные этапы относительно кратковременны.

Критерий  $k_{22}$  превосходит  $k_{23}$  по значимости, так как процесс шифрования и обмена сертификатами требует больше ресурсов, чем простой обмен пакетами на этапе установки соединения. Однако по сравнению с  $k_{24}$  роль  $k_{22}$  менее значительна, так как длительность загрузки данных является основным показателем скорости доставки данных.

Критерий  $k_{23}$  в свою очередь уступает всем остальным. Этот этап достаточно быстрый и не

создает узких мест в процессе доставки данных. В сравнении с  $k_{24}$  критерий  $k_{23}$  демонстрирует наименьшую значимость, так как передача данных, особенно при больших объемах, влияет на общее время гораздо сильнее.

Итоговая матрица парных сравнений (табл. 1) рассчитана на основе вышеизложенных рассуждений.

Наибольший вес был у скорости передачи данных, за ним следуют DNS, TLS и время соединения. Округлим веса до сотых, получим

$$w_{21} = 0,19, w_{22} = 0,1, w_{23} = 0,05, w_{24} = 0,66.$$

Для анализа частных критериев  $k_1$  также проведем попарное сравнение. Уменьшение времени работы скриптов оказывает прямое и комплексное влияние на производительность, так как охватывает обработку данных, рендеринг и скорость взаимодействия. Уменьшение количества запросов к базе данных снижает нагрузку на нее, но эффективность скриптов имеет больший приоритет, поскольку медленные скрипты могут нивелировать выгоду от оптимизации запросов. Быстродействие скриптов важнее оптимизации времени выполнения запросов, так как общее время работы скриптов охватывает больший спектр операций. Снижение объема данных страницы ускоряет передачу по сети, но это дает небольшой эффект, если основное время загрузки уходит на выполнение скриптов.

Уменьшение времени выполнения запросов к базе данных имеет более прямое влияние на производительность, так как медленные запросы могут вызвать значительные задержки даже при их небольшом количестве. Снижение количества запросов к базе данных важно для уменьшения нагрузки на сервер, особенно в системах с высокой активностью. Уменьшение объема данных страницы улучшает скорость передачи данных между сервером и клиентом, но избыточные запросы к базе данных оказывают большее влияние на общую производительность.

Уменьшение времени выполнения запросов к базе данных ускоряет обработку данных на сервере, минимизируя задержки в формирова-

нии страницы, что важно для работы с большими базами данных. Уменьшение объема данных страницы улучшает передачу данных по сети, но не компенсирует медленное выполнение запросов. Эти рассуждения легли в основу матрицы парных сравнений (табл. 2).

Наибольший вес был присвоен времени работы скриптов и времени выполнения запросов к базе данных — оба критерия оказывают прямое влияние на скорость формирования страницы. За ним следует количество запросов к базе данных — избыточное количество запросов может существенно замедлить работу системы. Наименьший вес получил объем созданной страницы. Округлив веса до сотых, получим

$$w_{11} = 0,39, w_{12} = 0,15, w_{13} = 0,39, w_{14} = 0,07.$$

Для критериев второго уровня иерархической модели также была составлена матрица парных сравнений. В ней критерий  $k_3$  оценен как более значимый по сравнению с  $k_2$ . Это объясняется тем, что методы оптимизации веб-страницы, такие как отложенная загрузка изображений или применение фасадов, способны оптимизировать загрузку страницы, уменьшая объем передаваемых данных. Такой подход снижает зависимость от характеристик передачи данных и компенсирует возможные задержки, делая влияние критерия  $k_3$  более значимым. Влияние остальных критериев было оценено как равнозначное, так как по отношению к общей производительности системы оно носит взаимосвязанный и сбалансированный характер. Такое распределение весов отражает сравнительную важность каждого критерия в контексте формирования страницы. Получившаяся матрица парных сравнений приведена в табл. 3.

Наибольший вес получил клиентский уровень, так как его оптимизация позволяет уменьшить объем передаваемых данных и компенсировать сетевые задержки. Серверный уровень уступает ему по значимости, но превосходит сетевой уровень, влияние которого оказалось наименьшим:

$$w_1 = 0,33, w_2 = 0,26, w_3 = 0,41.$$

■ **Таблица 1.** Парные сравнения частных критериев  $k_2$   
 ■ **Table 1.** Pairwise comparisons of partial criteria  $k_2$

	$k_{21}$	$k_{22}$	$k_{23}$	$k_{24}$
$k_{21}$	1	2	4	1/4
$k_{22}$	1/2	1	2	1/8
$k_{23}$	1/4	1/2	1	1/9
$k_{24}$	4	8	9	1

■ **Таблица 2.** Парные сравнения частных критериев  $k_1$   
 ■ **Table 2.** Pairwise comparisons of partial criteria  $k_1$

	$k_{11}$	$k_{12}$	$k_{13}$	$k_{14}$
$k_{11}$	1	3,00	1,00	5,00
$k_{12}$	0,33	1	0,33	3,00
$k_{13}$	1,00	3,00	1	5,00
$k_{14}$	0,20	0,33	0,20	1

■ Таблица 3. Парные сравнения критериев  
 ■ Table 3. Pairwise comparisons of criteria

	$k_1$	$k_2$	$k_3$
$k_1$	1	1,00	1,00
$k_2$	1,00	1	0,50
$k_3$	1,00	2,00	1



■ Рис. 3. Распределение весов критериев иерархической модели для уровней 2 и 3  
 ■ Fig. 3. Distribution of criteria weights of the hierarchical model for levels 2 and 3

Распределение весов критериев иерархической модели визуализировано на рис. 3. Подставив числовые коэффициенты, получим математическую формулу, описывающую иерархическую модель критериев:

$$K = 0,33 \cdot (0,39 \cdot k_{11} + 0,15 \cdot k_{12} + 0,39 \cdot k_{13} + 0,07 \cdot k_{14}) + 0,26 \cdot (0,19 \cdot k_{21} + 0,1 \cdot k_{22} + 0,05 \cdot k_{23} + 0,66 \cdot k_{24}) + 0,41 \cdot (0,1 \cdot k_{31} + 0,25 \cdot k_{32} + 0,3 \cdot k_{33} + 0,25 \cdot k_{34} + 0,1 \cdot k_{35}),$$

где значение каждого частного критерия  $k_{jt}$  рассчитывается как разность между значениями метрики до ( $b_{jt}$ ) и после ( $z_{jt}$ ) применения метода оптимизации. Если уменьшение численного значения метрики положительно влияет на скорость загрузки страницы (например, уменьшилось число запросов к базе данных),  $k_{jt} = b_{jt} - z_{jt}$ . Если положительный эффект достигается при увеличении метрики (например, возросла скорость передачи данных), то  $k_{jt} = z_{jt} - b_{jt}$ . Результат выражается в баллах от -100 до 100.

### Оценка методов по иерархической модели критериев

Рассмотрим применение иерархической модели критериев для поиска оптимальных методов оптимизации скорости загрузки веб-страницы на примере сайта chef2you.ru, разработанного авторами статьи для проведения эксперимента. Для оптимизации возьмем типовую страницу <https://chef2you.ru/retsept/lenivaia-pasta-s-miasom-i-ovoshhami>.

На основании рассмотренных работ [10–19] мы выделили 14 методов по улучшению скорости загрузки сайтов, применимых к нашему сайту: отложенная загрузка JavaScript-файлов, подключение JavaScript-файлов в конце HTML-кода; асинхронная загрузка JavaScript-файлов; минифицированный CSS- и JavaScript-код; сжатие медиафайлов (изображений, видео, аудио); критические CSS в начале страницы; не критические стили после блоков, для которых они написаны; разные по размеру изображения для разных размеров экрана; декларация размеров изображений через width и height; удаление устаревших или неиспользуемых html-тегов; отложенная загрузка медиаконтента («ленивая загрузка», фасады); локализация внешних ресурсов или предварительное подключение к внешним ресурсам; перемещение внешних файлов CSS и JavaScript размером до 1 КБ внутрь HTML; оптимизация шрифтов (формат, способ подключения). В рассмотренных материалах было предложено большее количество методов, но мы выбрали только те из них, которые могут быть применены к экспериментальному сайту chef2you.ru. Так, например, метод оптимизации, заключающийся в объединении нескольких маленьких изображений в одно большое, неприменим, так как на сайте chef2you.ru нет таких изображений.

Для измерения частных критериев  $k_1$  используется Middleware, измеряющий количество запросов к базе данных, их длительность и время генерации страницы (данные добавляются в html-код страницы). Вес страницы оценивается через консоль браузера. Для измерения частных критериев  $k_2$  используется служебная программа curl. Для измерения  $k_3$  используется Lighthouse.

При измерении частных критериев  $k_2$  следует учитывать, что скорость может отличаться в зависимости от состояния сети, к которой подключено устройство, проводившее проверку. Это показывают авторы статьи [20]. Ни один из 14 перечисленных методов оптимизации не влияет на критерий  $k_2$  и его частные критерии, поэтому в расчетах изменение состояния  $k_2$  всегда будет равно нулю.

Частный критерий  $k_{12}$  всегда равен нулю, так как количество запросов к базе данных не менялось при применении методов оптимизации

(всегда 34 запроса). Критерии  $k_{13}$  и  $k_{11}$  равны нулю, так как изменения в их метриках являются статистически незначимыми. Это было доказано путем проведения дисперсионного анализа (ANOVA) [21].

Для измерения изменения метрики объема данных страницы воспользуемся формулой арктангенсной нормализации с учетом значимого изменения, выведенной авторами статьи:

$$k_{14} = \frac{\arctan((b_{14} - z_{14}) \cdot d)}{\pi / 2} \cdot 100.$$

Формула основана на способе нормализации функции через арктангенс, описанном в статье [22]. Положительным влиянием на скорость загрузки страницы считается уменьшение размера страницы. Чтобы отразить это в формуле, из размера страницы до применения метода оптимизации вычитается вес страницы после применения метода оптимизации ( $b_{14} - z_{14}$ ) в килобайтах. Функция позволяет управлять степенью значимости изменения метрики ( $b_{14} - z_{14}$ ) с помощью параметра  $d$ , сохраняя при этом свойства монотонности и ограниченности функции. Для приведения диапазона значений функции  $[-\pi/2; +\pi/2]$  к диапазону  $[-1; +1]$  результат  $\arctan((b_{14} - z_{14}) \cdot d)$  делится на  $\pi/2$ . Так как все критерии в рассматриваемой модели лежат в диапазоне  $[-100; +100]$ , применяется масштабирование, результат умножается на 100.

Таким образом, в нашей формуле  $d$  является коэффициентом значимости, равным 0,000617, при котором изменение веса страницы на 10 МБ даст оценку влияния 90 баллов из 100. Такая формула позволит измерить изменение веса страницы в диапазоне от 0 до 100. Например, исходный размер оптимизируемой страницы равен 13,60 КБ, после внедрения первого метода оптимизации размер страницы стал 14,80 КБ. Изменение размера составило 1,20 КБ, что дает -0,05 балла для МКА.

Рассмотрим влияние методов оптимизации веб-страницы на скорость ее загрузки. Первый метод заключается в отложенной загрузке скриптов. Для этого на страницу добавляется код, организующий очередь из задач загрузки js скриптов и выполнения js функций, которые начинают исполняться после определенного момента (в данном случае после полной загрузки страницы без скриптов). Эффективность описанного метода заметна при разделении массива кода с высоким временем исполнения. Однако тестовый сайт не имел таких проблем изначально, поэтому по данным Lighthouse скорость загрузки страницы увеличилась незначительно, с 47,67 (табл. 4) до 48,67 (табл. 5).

Подставим данные измерения в иерархическую модель критериев. Как было упомяну-

■ **Таблица 4.** Метрики для параметра  $k_3$  и его частных параметров до применения первого метода оптимизации на настольных устройствах

■ **Table 4.** Metrics for parameter  $k_3$  and its private parameters before applying the first optimization method on desktop devices

Значение	Номер теста			Среднее
	1	2	3	
$b_{31}$ для $k_{31}$ (FCP)	98	96	96	96,67
$b_{32}$ для $k_{32}$ (LCP)	6	3	6	5
$b_{33}$ для $k_{33}$ (TBT)	7	15	7	9,67
$b_{34}$ для $k_{34}$ (CLS)	99	99	100	99,33
$b_{35}$ для $k_{35}$ (SI)	87	81	96	88
Оценка Lighthouse	47	48	48	47,67

■ **Таблица 5.** Метрики для параметра  $k_3$  и его частных параметров после применения первого метода оптимизации на настольных устройствах

■ **Table 5.** Metrics for parameter  $k_3$  and its private parameters after applying the first optimization method on desktop devices

Значение	Номер теста			Среднее
	1	2	3	
$z_{31}$ для $k_{31}$ (FCP)	97	90	98	95
$z_{32}$ для $k_{32}$ (LCP)	4	4	4	4
$z_{33}$ для $k_{33}$ (TBT)	7	15	21	14,33
$z_{34}$ для $k_{34}$ (CLS)	99	99	99	99
$z_{35}$ для $k_{35}$ (SI)	97	91	95	94,33
Оценка Lighthouse	47	48	51	48,67

то ранее,  $k_2 = 0$ , так как все частные критерии  $k_{21} - k_{24}$  равны нулю. Критерий  $k_{14}$  был вычислен с помощью формулы арктангенсной нормализации с учетом значимого изменения (нужно было выразить изменение размера файла в килобайтах баллами от -100 до 100). Найти частные критерии  $k_{31} - k_{35}$  можно путем вычитания из значеный метрик после применения метода оптимизации ( $z_{jt}$ ) значений метрик до применения метода оптимизации ( $b_{jt}$ ). Lighthouse предоставляет значения метрик, выраженные в баллах от 0 до 100, поэтому приводить его оценку к 100-балльной шкале не требуется. Позитивным влиянием считается увеличение значений метрик по Lighthouse после применения метода оптимизации:

$$k_{3t} = z_{3t} - b_{3t};$$

$$k_{31} = 95 - 96,67 = -1,67;$$

$$k_{32} = 4 - 5 = -1;$$

$$k_{33} = 14,33 - 9,67 = 4,66;$$

$$k_{34} = 99 - 99,33 = -0,33;$$

$$k_{35} = 94,33 - 88 = 6,33;$$

$$k_1 = 0,39 \cdot 0 + 0,15 \cdot 0 + 0,39 \cdot 0 + 0,07 \cdot (-0,05) = -0,04;$$

$$k_3 = 0,1 \cdot (-1,67) + 0,25 \cdot (-1) + 0,3 \cdot 4,66 + 0,25 \cdot (-0,33) + 0,1 \cdot 6,33 = 1,53;$$

$$K = 0,33 \cdot (-0,04) + 0,26 \cdot 0 + 0,41 \cdot 1,53 = 0,6.$$

По аналогии были оценены остальные методы оптимизации для мобильных и настольных устройств (рис. 4).

Метод отложенной загрузки медиаконтента («ленивая загрузка», фасады) показал наибольшую эффективность по сравнению с другими методами оптимизации. После его внесения можно повторно рассчитать оценку для оставшихся методов оптимизации. Метод, показывающий наибольшую эффективность, можно применять для оптимизации скорости загрузки страницы.

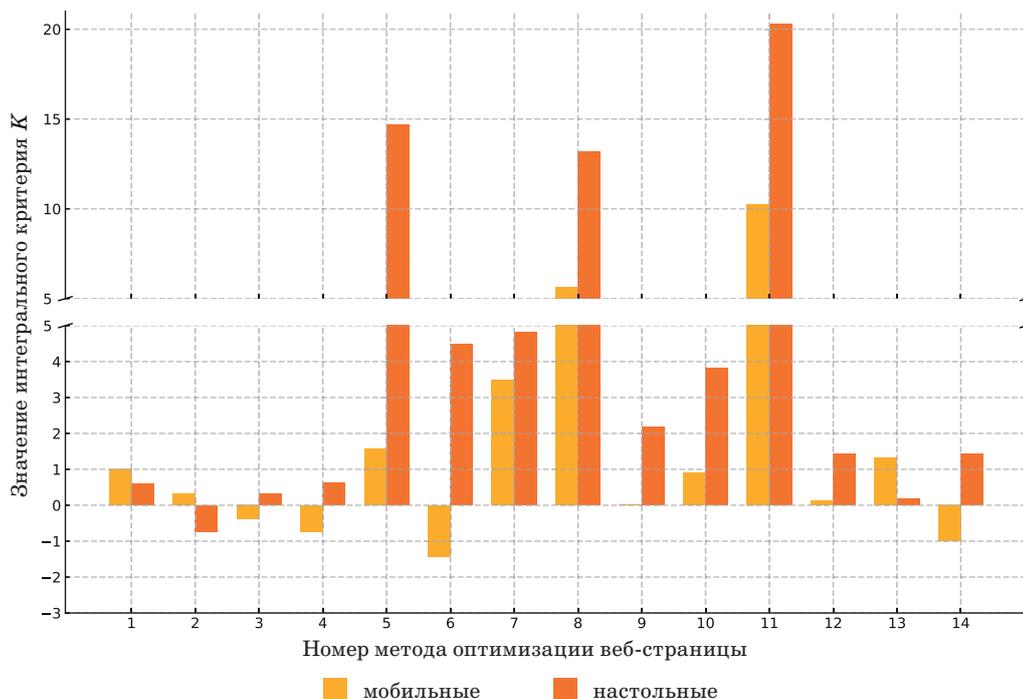
### Влияние методов на поведение пользователей

Для проверки влияния предложенных методов оптимизации загрузки веб-страниц был разработан единый веб-ресурс на базе Laravel,

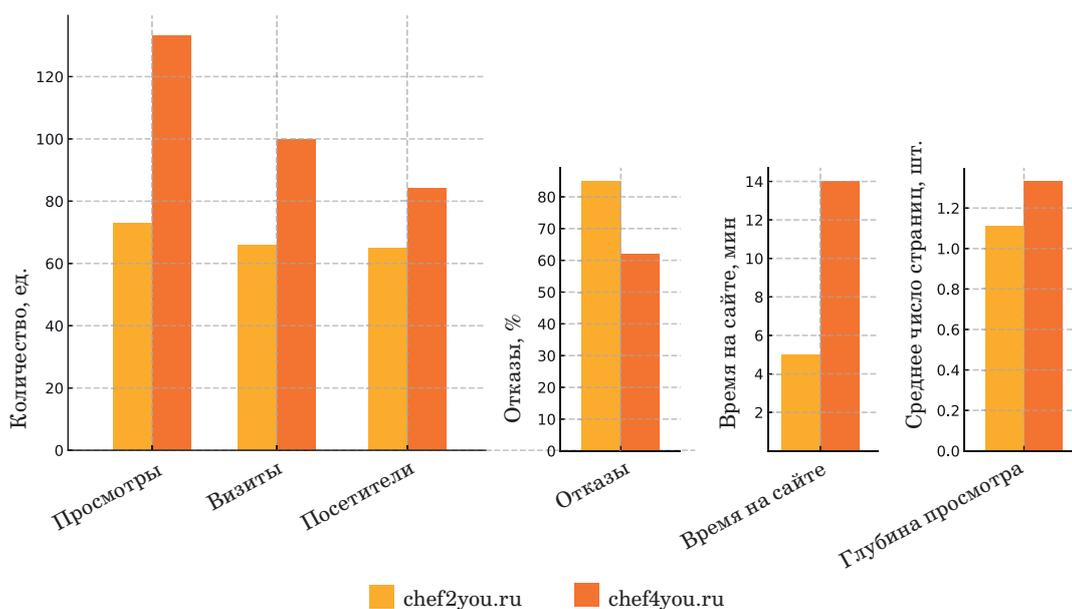
включающий два отдельных шаблона. Каждый шаблон подключен к своему набору данных (текста и изображений, чтобы избежать дублирования контента в глазах поисковых систем) и размещен на отдельном домене. Таким образом, с точки зрения пользователей и поисковых систем это два независимых сайта. Оба шаблона визуально идентичны и содержат одинаковую по смыслу информацию – страницы с рецептами блюд.

Первый шаблон реализован без учета методов оптимизации веб-страниц и подключен к домену chef2you.ru (на его странице в предыдущем разделе были произведены замеры эффективности методов оптимизации). Второй шаблон был сделан позже и подключен к домену chef4you.ru. Он реализует методы оптимизации, набравшие больше 10 баллов положительного влияния на скорость загрузки страницы (отложенная загрузка медиаконтента, сжатие медиафайлов, разные по размеру изображения для разных размеров экрана) на основании анализа страницы сайта chef2you.ru.

В процессе разработки сайты были закрыты от индексации для поисковых систем через правила в файле robots.txt. После окончания разработки сайты были одновременно открыты для индексации, добавлены в Google Search Console и Яндекс Вебмастер. На сайты были установлены скрипты Яндекс Метрики – это инструмент



■ **Рис. 4.** Результаты оценки влияния методов оптимизации на скорость загрузки веб-страницы  
 ■ **Fig. 4.** Results of the evaluation of the impact of optimization methods on the loading speed of a web page



■ **Рис. 5.** Статистика посещаемости сайтов  
 ■ **Fig. 5.** Website traffic statistics

веб-аналитики, позволяющий отслеживать множество показателей сайта, среди которых статистика визитов. По итогу наблюдений за сайтами в течение трех месяцев (четвертый квартал 2024 года) были получены данные, представленные на рис. 5.

По статистике видно, что сайт chef4you.ru превосходит сайт chef2you.ru по всем основным показателям. Это доказывает влияние скорости загрузки веб-страницы на поведение ботов и пользователей. Исходя из того, что сайты отличаются только технической реализацией, можно сделать вывод, что пользователи и боты предпочитают выбирать сайты, соблюдающие советы по оптимизации скорости загрузки.

### Заключение

Разработанная иерархическая модель критериев может быть полезна как для оптимизации существующих веб-ресурсов, так и для планирования новых проектов. Благодаря такой модели

разработчики, аналитики и владельцы сайтов могут принимать более обоснованные решения, направленные на повышение скорости загрузки страниц, улучшение пользовательского опыта и, как следствие, достижение лучших результатов в удержании посетителей на целевом ресурсе.

Следует отметить, что не все методы оптимизации применимы к различным типам сайтов. Например, сайты, созданные с помощью визуальных конструкторов, ограничены в возможностях прямой настройки серверной части и в изменении кода, генерирующего страницу. Однако модель сохраняет свою актуальность и в этих случаях. Она позволяет использовать и оценивать эффективность тех методов, которые доступны на выбранной платформе. В случае недоступности некоторых метрик (например, времени запросов) можно воспользоваться приближенными оценками или принять нулевые значения, сохранив возможность корректного сравнения различных методов. Это делает модель универсальным инструментом анализа в широком спектре веб-разработки.

### Литература

1. Долгова Н. Э., Долгов И. М. Эволюция Web3 и интернет-реклама: от зарождения Интернета до современных трендов и экономических последствий. *Экономика и управление: проблемы, решения*, 2024, т. 1, № 9(150), с. 149–157. doi:10.36871/ek.чр.p.r.2024.09.01.015, EDN: CWGKUV

2. Драч В. Е., Ильичев В. Ю. Эффективные способы повышения рейтинга сайта в Google. *Системный администратор*, 2021, № 10(227), с. 84–88. EDN: FYPPKO

3. Китаев Е. Л., Скорнякова Р. Ю. Использование микроразметок для добавления в контент веб-страницы данных внешних ресурсов. *Электронные библиотеки*, 2020, т. 23, № 3, с. 494–513. <https://>

- elbib.ru/article/view/597 (дата обращения: 26.03.2025). doi:10.26907/1562-5419-2020-23-3-494-513, EDN: ORWZXC
4. **Полухин Н. В., Эккерт Н. В., Козлов В. В.** Информационные предпочтения пациентов в контексте коммуникации на веб-сайтах медицинских организаций. *Siberian Journal of Life Sciences and Agriculture*, 2021, т. 13, № 5, с. 226–246. doi:10.12731/2658-6649-2021-13-5-226-246, EDN: SIDPPS
  5. **Özkan B., Özceylan E., Kabak M., Dağdeviren M.** Evaluating the websites of academic departments through SEO criteria: A hesitant fuzzy linguistic MCDM approach. *Artificial Intelligence Review*, 2020, vol. 53, no. 2, pp. 875–905. doi:10.1007/s10462-019-09681-z
  6. **Жуков С. В., Ковалева О. А., Ковалев С. В.** Концептуальная модель загрузки веб-страницы. *Информационные технологии и вычислительные системы*, 2024, № 2, с. 26–36. doi:10.14357/20718632240203, EDN: LWCJNC
  7. **Баранова Е. М., Баранов А. Н., Борзенкова С. Ю., Васин К. И., Пересяблов В. С.** Современные тенденции защиты баз данных веб-приложений от SQL инъекций. *Известия Тульского государственного университета. Технические науки*, 2023, № 12, с. 492–495. doi:10.24412/2071-6168-2023-12-492-493, EDN: GLVPBS
  8. **Лубенцова Е. В., Ожогова Е. В., Лубенцов В. Ф., Шахрай Е. А., Масютина Г. В.** Метод согласованности матриц парных сравнений на основе компонент их максимальных собственных чисел. *Современные наукоемкие технологии*, 2020, № 7, с. 78–83. doi:10.17513/snt.38138, EDN: DJHVKM
  9. **Рычаго М. Е., Хорошева А. В.** Некоторые особенности применения метода парных сравнений при динамическом изменении количества альтернатив в иерархической структуре данных. *Перспективы науки*, 2022, № 1(148), с. 14–18. EDN: YDEDZY
  10. **Жуков С. В., Ковалева О. А., Ковалев С. В.** Анализ и разработка алгоритмов, оптимизирующих компоновку html документа для браузеров. *Информационные технологии и вычислительные системы*, 2023, № 3, с. 114–126. doi:10.14357/20718632230312, EDN: VITVWW
  11. **Максимов Я. А., Мартышкин А. И.** Возможности оптимизации производительности веб-приложений. *Современные информационные технологии*, 2024, № 39(39), с. 189–192. EDN: INPFTW
  12. **Галимуллин Н. Р., Гузуева Э. Р.** Разработка интерактивных веб-приложений с использованием современных javascript фреймворков. *Экономика и управление: проблемы, решения*, 2024, т. 7, № 11(152) с. 167–178. doi:10.36871/ek.up.p.r.2024.11.07.017, EDN: BQEUGL
  13. **Ильин А. Ю., Плотников С. Б.** Использование предиктивного анализа пользовательского поведения для повышения скорости реактивной загрузки клиентской части веб-приложения. *Международный журнал гуманитарных и естественных наук*, 2025, № 1-3(100), с. 140–143. doi:10.24412/2500-1000-2025-1-3-140-143, EDN: OUIRVA
  14. **Гридин В. Н., Анисимов В. И., Васильев С. А.** Методы повышения производительности современных веб-приложений. *Известия ЮФУ. Технические науки*, 2020, № 2(212), с. 193–200. doi:10.18522/2311-3103-2020-2-193-200, EDN: BGUPQL
  15. **Чимидов Д. Б., Эляев О. А., Эренженова Э. Э., Лиджиева З. А., Цебеков С. И., Кирсанова В. Л., Баркуева Е. В.** Анализ и оптимизация производительности веб-приложений. *Экономика и предпринимательство*, 2024, № 8(169), с. 1139–1142. doi:10.34925/EIP.2024.169.8.213, EDN: EGFSLL
  16. **Маскинскова У. В., Ямашкин С. А., Медянкин И. В., Сидорин А. В., Ермолаев М. Е.** Разработка и оптимизация веб-ресурса с функциями социальной сети. Описание ролей с анализом данных и учетом показаний мониторинга автоматизированной системы. *Научно-технический вестник Поволжья*, 2024, № 1, с. 96–98. EDN: NGPCVH
  17. **Присяжный А. О.** Оптимизация кроссбраузерной эффективности высокопосещаемых веб-приложений. *Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки*, 2024, № 7-2, с. 163–167. doi:10.37882/2223-2966.2024.7-2.29, EDN: PBUWCS
  18. **Торопкин Р. А., Зиновьев Я. В., Рассказов Н. С., Митрохин М. А.** Технологии оптимизации работы сайта на примере аналитической системы публикационной активности Пензенского государственного университета. *Модели, системы, сети в экономике, технике, природе и обществе*, 2020, № 4(36), с. 71–78. doi:10.21685/2227-8486-2020-4-7, EDN: WMRUQA
  19. **Горячкин Б. С., Ханмурзин Т. И.** Повышение эффективности работы с веб-ресурсом за счет инструментария системного программиста. *Динамика сложных систем – XXI век*, 2022, т. 16, № 3, с. 26–39. doi:10.18127/j19997493-202203-03, EDN: OVVCQU
  20. **Хорошко М. Б., Джавад Х. А.** Изменение скорости подключения к интернету в зависимости от количества сетей на одном WI-FI канале. *Инженерный вестник Дона*, 2020, № 8(68), с. 290–298. EDN: GJRTDM
  21. **Дисперсионный анализ (ANOVA).** <https://habr.com/ru/companies/otus/articles/734258/> (дата обращения: 08.03.2025).
  22. **Старовойтов В. В., Голуб Ю. И.** Нормализация данных в машинном обучении. *Информатика*, 2021, т. 18, № 3, с. 83–96. doi:10.37661/1816-0301-2021-18-3-83-96, EDN: JKANKM

UDC 004.052.42

doi:10.31799/1684-8853-2025-4-45-57

EDN: LPNYCL

**Multi-criteria analysis of web page optimization methods and their impact on search engine rankings**S. V. Zhukov<sup>a</sup>, Post-Graduate Student, orcid.org/0009-0003-8954-8565O. A. Kovaleva<sup>a</sup>, Dr. Sc., Tech., Associate Professor, orcid.org/0000-0003-0735-6205, solomina-oa@yandex.ruS. V. Kovalev<sup>a</sup>, Dr. Sc., Tech., Associate Professor, orcid.org/0000-0002-5961-7561<sup>a</sup>Derzhavin Tambov State University, 33, Internatsionalnaya St., 392000, Tambov, Russian Federation

**Introduction:** Web page loading speed is one of the critical factors influencing website rankings in search engines. Slow loading negatively impacts user experience; consequently, it may lead to losing potential customers and decreasing conversions. Moreover, in an increasingly competitive environment, speed optimization becomes an essential part of effective website development, directly affecting a site's success online. **Purpose:** To perform a multi-criteria analysis for identifying the most effective web page speed optimization methods for a specific website. **Results:** We develop a hierarchical model of criteria to evaluate the effectiveness of various optimization methods, incorporating server-side, network, and client-side indicators for webpage loading speed. In addition, we analyze webpage optimization methods using a specific website as an example. The results demonstrate that implementing deferred media content loading (lazy loading) provides the most significant improvement. Certain optimization methods show no meaningful impact on the server-side and network performance criteria. Based on these findings, we have created a second website employing the most effective webpage optimization techniques. As a result of a comparative experiment, the optimized website demonstrates substantial improvements across all key metrics, such as the number of visits, duration of visits, page views per session, and a decrease in bounce rate. **Practical relevance:** The benefits of this study and the proposed model enable developers to make informed decisions when selecting the most effective methods to optimize web page loading speed.

**Keywords** – web page loading speed, web page optimization, multi-criteria analysis, paired comparison method, search engine optimization, behavioral factors, web development.

**For citation:** Zhukov S. V., Kovaleva O. A., Kovalev S. V. Multi-criteria analysis of web page optimization methods and their impact on search engine rankings. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 45–57 (In Russian). doi:10.31799/1684-8853-2025-4-45-57, EDN: LPNYCL

**References**

- Dolgova N. E., Dolgov I. M. The evolution of Web3 and internet advertising: From the birth of the Internet to modern trends and economic impacts. *Ekonomika i upravlenie: problemy, resheniya*, 2024, vol. 1, no. 9(150), pp. 149–157 (In Russian). doi:10.36871/ek.up.p.r.2024.09.01.015, EDN: CWGKUV
- Drach V. E., Ilichev V. Y. Development of software tools to enlarge images using their fractal properties. *Sistemnyy administrator*, 2021, no. 10(227), pp. 84–88 (In Russian). EDN: FYPPKO
- Kitaev E. L., Skorniyakova R. Yu. Leveraging semantic markups for incorporating external resources data to the content of a web page. *Russian Digital Libraries Journal*, 2020, vol. 23, no. 3, pp. 494–513. Available at: <https://elbib.ru/article/view/597> (accessed 26 March 2025) (In Russian). doi:10.26907/1562-5419-2020-23-3-494-513, EDN: ORWZXC
- Polukhin N. V., Ekkert N. V., Kozlov V. V. Patient perspective on healthcare facilities websites communication elements. *Siberian Journal of Life Sciences and Agriculture*, 2021, vol. 13, no. 5, pp. 226–246 (In Russian). doi:10.12731/2658-6649-2021-13-5-226-246, EDN: SIDPPS
- Ozkan B., Özceylan E., Kabak M., Dağdeviren M. Evaluating the websites of academic departments through SEO criteria: A hesitant fuzzy linguistic MCDM approach. *Artificial Intelligence Review*, 2020, vol. 53, no. 2, pp. 875–905. doi:10.1007/s10462-019-09681-z
- Zhukov S. V., Kovaleva O. A., Kovalev S. V. Conceptual model of web page loading. *Journal of Information Technologies and Computing Systems*, 2024, no. 2, pp. 26–36 (In Russian). doi:10.14357/20718632240203, EDN: LWCJNC
- Baranova E. M., Baranov A. N., Borzenkova S. Yu., Vasin K. I., Perezyabov V. S. Current trends in database protection web applications from SQL injection. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskiye nauki*, 2023, no. 12, pp. 492–495 (In Russian). doi:10.24412/2071-6168-2023-12-492-493, EDN: GLVPBS
- Lubentsova E. V., Ozhogova E. V., Lubentsov V. F., Shakh-ray E. A., Masyutina G. V. Method of matching paired comparison matrices based on components of their maximum eigenfrequency numbers. *Sovremennyye naukoemykiye tekhnologii*, 2020, no. 7, pp. 78–83 (In Russian). doi:10.17513/snt.38138, EDN: DJHVKM
- Rychago M. E., Khorosheva A. V. Some aspects of the application of the method of paired comparisons with a dynamic change in the number of alternatives in the hierarchical data structure. *Science Prospects*, 2022, no. 1(148), pp. 14–18 (In Russian). EDN: YDEDZY
- Zhukov S. V., Kovaleva O. A., Kovalev S. V. Analysis and development of algorithms that optimize the layout of the html document for browsers. *Journal of Information Technologies and Computing*, 2023, no. 3, pp. 114–126 (In Russian). doi:10.14357/20718632230312, EDN: VITVWW
- Maximov YA. A., Martyshekin A. I. Possibilities of web applications performance optimisation. *Sovremennyye informatsionnyye tekhnologii*, 2024, no. 39(39), pp. 189–192 (In Russian). EDN: INPFTW
- Galimullin N. R., Guzueva E. R. Developing interactive web applications using modern javascript frameworks. *Ekonomika i upravlenie: problemy, resheniya*, 2024, vol. 7, no. 11, pp. 167–178 (In Russian). doi:10.36871/ek.up.p.r.2024.11.07.017, EDN: BQEUGL
- Ilyin A. Y., Plotnikov S. B. Using predictive analysis of user behavior to increase the speed of reactive loading of the client part of the web application. *International Journal of Humanities and Natural Sciences*, 2025, no. 1-3(100), pp. 140–143 (In Russian). doi:10.24412/2500-1000-2025-1-3-140-143, EDN: OUIRVA
- Gridin V. N., Anisimov V. I., Vasilev S. A. Methods of improving performance of modern web-applications. *Izvestiya SFedU. Engineering Sciences*, 2020, no. 2(212), pp. 193–200 (In Russian). doi:10.18522/2311-3103-2020-2-193-200, EDN: BGUPQL
- Chimidov D. B., Elyayev O. A., Erenzhenova E. E., Lidzheva Z. A., Tsebekov S. I., Kirsanova V. L., Barkueva E. V. Analysis and optimization of web application performance. *Ekonomika i predprinimatel'stvo*, 2024, no. 8(169), pp. 1139–1142 (In Russian). doi:10.34925/EIP.2024.169.8.213, EDN: EGFSL
- Maskinskova U. V., Yamashkin S. A., Medyankin I. V., Sidorin A. V., Ermolaev M. E. Development and optimization of a web resource with social network functions. Description of roles with data analysis and taken into account of automated system monitoring indications. *Scientific and Technical Volga region Bulletin*, 2024, no. 1, pp. 96–98 (In Russian). EDN: NGPCVH
- Prisyazhnyy A. O. Optimization of cross-browser efficiency of highly visited web applications. *Sovremennaya nauka: aktualnyye problemy teorii i praktiki. Seriya: Yestestvennyye i tekhnicheskiye nauki*, 2024, no. 7-2, pp. 163–167 (In Russian). doi:10.37882/2223-2966.2024.7-2.29, EDN: PBUWCS
- Toropkin R. A., Zinov'yev YA. V., Rasskazov N. S., Mitrokhin M. A. Site optimization technologies on the example of the analytical system of publishing activity of Penza State Uni-

- versity. *Models, Systems, Networks in Economics, Engineering, Nature and Society*, 2020, no. 4(36), pp. 71–78 (In Russian). doi:10.21685/2227-8486-2020-4-7, EDN: WMRUQA
19. Goryachkin B. S., Khanmurzin T. I. Improving the efficiency of working with a web resource due to the tools of a system programmer. *Dynamics of Complex Systems – XXI Century*, 2022, vol. 16, no. 3, pp. 26–39 (In Russian). doi:10.18127/j19997493-202203-03, EDN: OVVCQU
20. Horohko M. B., Djavad H. A. Changing the internet connection speed depending on the number of networks on a single WI-FI channel. *Engineering Journal of Don*, 2020, no. 8(68), pp. 290–298 (In Russian). EDN: GJRTDM
21. *Dispersionnyy analiz (ANOVA)* [Analysis of Variance (ANOVA)] Available at: <https://habr.com/ru/companies/otus/articles/734258/> (accessed 8 March 2025).
22. Starovoitov V. V., Golub YU. I. Data normalization in machine learning. *Informatics*, 2021, vol. 18, no. 3, pp. 83–96 (In Russian). doi:10.37661/1816-0301-2021-18-3-83-96, EDN: JKAHKM
- 

---

### ПАМЯТКА ДЛЯ АВТОРОВ

*Поступающие в редакцию статьи проходят обязательное рецензирование.*

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail ([ius.spb@gmail.com](mailto:ius.spb@gmail.com)).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию.

*Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.*

---



## Метод определения уязвимостей программного кода на основе кластерного анализа и контекстной адаптации больших языковых моделей

Р. Н. Бакеев<sup>а</sup>, руководитель центра, [orcid.org/0009-0008-3149-3836](https://orcid.org/0009-0008-3149-3836)

В. Н. Кузьмин<sup>б</sup>, доктор воен. наук, ведущий научный сотрудник, [orcid.org/0000-0002-6411-4336](https://orcid.org/0000-0002-6411-4336), [vka@mil.ru](mailto:vka@mil.ru)

А. Б. Менисов<sup>б</sup>, доктор техн. наук, старший преподаватель, [orcid.org/0000-0002-9955-2694](https://orcid.org/0000-0002-9955-2694)

Т. Р. Сабиров<sup>б</sup>, канд. техн. наук, старший преподаватель, [orcid.org/0000-0002-6807-2954](https://orcid.org/0000-0002-6807-2954)

<sup>а</sup>Фонд перспективных исследований, Бережковская наб., 22, стр. 3, Москва, 121059, РФ

<sup>б</sup>Военно-космическая академия им. А. Ф. Можайского, Ждановская наб., 13, Санкт-Петербург, 197198, РФ

**Введение:** обнаружение уязвимостей в исходном коде остается одной из приоритетных задач в области обеспечения кибербезопасности. Классические методы анализа кода часто не учитывают контекст исполнения и слабо масштабируются при росте объема программ. В условиях сложных архитектур и неполных аннотаций данных требуется контекстно-адаптированный подход, способный выявлять уязвимости на основе семантического и структурного анализа. **Цель:** разработать метод обнаружения уязвимостей в программном коде с использованием кластерного анализа и контекстной адаптации больших языковых моделей, способных учитывать не только синтаксис, но и семантическую структуру программ. **Методы:** разработанный подход сочетает кластеризацию кодовых сегментов с применением предварительно дообученных языковых моделей, адаптированных к программному коду. Для повышения результативности используется выделение признаков уязвимости, включающих как непосредственно уязвимый фрагмент, так и его контекст – управляющие конструкции, переменные, вызовы функций. Обучение и оценка проводились на размеченных открытых датасетах с использованием предобученных больших языковых моделей. **Результаты:** метод позволяет автоматически группировать фрагменты кода по структурному сходству, после чего осуществляется семантический анализ с помощью больших языковых моделей, способных распознавать шаблоны уязвимостей. Эксперименты показали, что включение контекстной информации существенно повышает эффективность определения уязвимостей в исходном коде. На датасетах BigVul и CVEfixes предложенный метод достиг точности до 78 % и полноты 82 %, что на 9–12 % выше по сравнению с существующими решениями. Метод демонстрирует устойчивость к синтаксическим вариациям и может быть использован для анализа ранее неразмеченного кода. **Практическая значимость:** метод применим в системах автоматического анализа исходного кода и способен значительно сократить затраты на ручной аудит, особенно в условиях анализа больших кодовых баз. Также он может использоваться в образовательных и исследовательских целях для анализа паттернов уязвимостей. **Обсуждение:** результаты подтверждают эффективность использования больших языковых моделей в задачах анализа безопасности программ. Перспективным направлением является расширение метода на другие языки программирования, а также исследование гибридных подходов с участием графовых нейросетей. Открытым остается вопрос аргументации решений модели и автоматического объяснения причин классификации сегмента как уязвимого.

**Ключевые слова** – информационная безопасность, обнаружение уязвимостей, большие языковые модели, кластерный анализ, машинное обучение.

**Для цитирования:** Бакеев Р. Н., Кузьмин В. Н., Менисов А. Б., Сабиров Т. Р. Метод определения уязвимостей программного кода на основе кластерного анализа и контекстной адаптации больших языковых моделей. *Информационно-управляющие системы*, 2025, № 4, с. 58–70. doi:10.31799/1684-8853-2025-4-58-70, EDN: XJDOLW

**For citation:** Bakeev R. N., Kuzmin V. N., Menisov A. B., Sabirov T. R. Method for identifying software code vulnerabilities based on cluster analysis and contextual adaptation of large language models. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 58–70 (In Russian). doi:10.31799/1684-8853-2025-4-58-70, EDN: XJDOLW

### Введение

Уязвимости программного обеспечения требуют значительных усилий по их обнаружению и устранению [1], так как представляют существенные проблемы в информационной безопасности [2]. Они имеют критическое значение, поскольку не устраненные уязвимости могут потенциально привести к скомпрометированной информационной инфраструктуре и огромным экономическим потерям [3]. Точное и быстрое выявление, а самое главное, обоснованное опре-

деление уязвимостей имеет решающее значение для снижения рисков [4] и привлекает значительное внимание со стороны ученых, разработчиков информационных систем и специалистов информационной безопасности [5, 6].

Обнаружение уязвимостей программного кода критично на всех этапах разработки безопасного программного обеспечения и в основном базируется на положениях динамических и статических подходов [7]. Динамические методы трудозатратны и направлены на выявление уязвимостей путем выполнения кода и наблюдения

за выводом или внутренними состояниями программы [8]. При реализации статического подхода анализ кода проводят без его выполнения, используя анализ синтаксиса, потока данных и управления. Инструменты статического тестирования безопасности приложений популярны благодаря своей низкой стоимости, быстрой работе и способности находить ошибки без запуска программы. Однако эти методы имеют главный недостаток — высокую долю ложноположительных результатов.

Применение моделей машинного обучения преимущественно относится к статическому анализу, который в последние годы стал ключевым направлением исследований [9]. Помимо глубокого обучения, недавние достижения в области больших языковых моделей (БЯМ) значительно повлияли на обнаружение уязвимостей программного обеспечения. Продвинутое методы, такие как дополненный поиск [10], контекстное обучение на примерах [11] и тонкая настройка моделей машинного обучения [12], уже использовались для обнаружения уязвимостей [7]. Вводя цепочки инструкций (промтов), основанные на семантической структуре кода, БЯМ могут достичь более высокой результативности обнаружения и обладают способностью рассуждения об уязвимостях. Хотя был достигнут значительный прогресс в применении методов глубокого обучения для обнаружения уязвимостей [13–15], эффективное использование этих методов для выявления и объяснения уязвимостей остается сложной проблемой. В настоящее время лишь ограниченное количество исследований фокусируется на объяснительных возможностях моделей, основанных на глубоком обучении, в отношении уязвимостей [16].

### Анализ научно-технических реализаций

Большие языковые модели в последние годы находят широкое применение в различных сферах экономики, включая автоматизацию бизнеса [17], анализ данных [18] и их безопасность [19]. В области информационной безопасности они используются для автоматизированного анализа угроз [20], распознавания вредоносных программ и обнаружения уязвимостей исходного кода [21]. Решения, основанные на БЯМ, продемонстрировали свою эффективность для обнаружения уязвимостей программного обеспечения [22, 23]. Применяемые БЯМ — это модели машинного обучения, включающие большое количество параметров и предварительно обученные на огромных объемах исходного кода, текста и других типах данных. В последнее время наблю-

дается значительное увеличение размера обучающих данных и параметров БЯМ [24]. Однако эффективность их применения в этих задачах зависит от способности адаптации к специфическим контекстам программирования и анализа безопасности. Даже современные БЯМ, такие как GPT-4 [25] и Code Llama [26], демонстрируют высокую точность при анализе программного кода, но требуют адаптации к конкретным доменам и типам уязвимостей.

В области развития БЯМ активно развиваются подходы, направленные на раскрытие внутренней логики вывода [27, 28] и объяснения классификационных решений в программном коде через контекстные зависимости переменных и вызовов [29].

Значительным достижением в обнаружении уязвимостей на основе БЯМ является эмуляция процессов экспертного рассуждения человека [30], т. е. БЯМ следует пошаговой схеме при анализе кода на наличие уязвимостей. Этот процесс гарантирует, что модели сначала генерируют цепочку рассуждений, прежде чем принимать решения о потенциальных уязвимостях. Дополнительно, используя контекстное обучение с несколькими примерами, реализуя возможности БЯМ, можно повысить результативность обнаружения уязвимостей, совершенствуя способность модели объяснять признаки для обнаружения, делая результаты более аргументированными и заслуживающими доверия. Пошаговое рассуждение также улучшает контекстное понимание фрагментов кода. Вместо того чтобы рассматривать код как изолированные сегменты, БЯМ способны анализировать взаимосвязи между различными компонентами и повышать контекст уязвимостей [31]. Это целостное понимание имеет решающее значение для обнаружения сложных уязвимостей, которые охватывают несколько функций или модулей.

Несмотря на значительные достижения в методологиях обнаружения уязвимостей, сохраняется ряд трудностей в обучении и оценивании этих систем.

1. Эффективность систем обнаружения уязвимостей, под которой в данной работе понимается совокупность таких метрик, как точность (precision), полнота (recall), F1-мера и уровень ложноположительных срабатываний (false positive rate), существенно зависит от характеристик используемых обучающих и тестовых наборов данных [32, 33].

Доказано, что даже высокоточные алгоритмы значительно теряют в производительности при обучении на плохо размеченных наборах данных [34]. Поэтому при разработке эффективных методов выявления уязвимостей необходим особый акцент на отбор, препроцессинг и валидацию

исходных датасетов. Эти проблемы с качеством данных обуславливают важность адекватного оценивания текущего состояния практики обнаружения уязвимостей и подчеркивают необходимость улучшать стратегии проверки наборов данных.

2. Многие открытые наборы данных в предметной области обнаружения уязвимостей имеют значительный дисбаланс классов, с гораздо большим количеством отрицательных примеров (безопасный код), чем положительных (уязвимый код). Устранение этого дисбаланса требует специализированных подходов для повышения эффективности обнаружения уязвимостей в программном коде от алгоритмических решений до дополнения наборов данных. Однако существуют и специализированные датасеты, включающие преимущественно или исключительно уязвимые прецеденты, например `gitbug-java` [35].

Поскольку разнообразие уязвимостей будет только возрастать, привлекательной и перспективной для постоянных исследований и разработок будет задача совершенствования методов обнаружения уязвимостей. Таким образом, необходимо сосредоточить усилия на устранении недостатков качества наборов данных, обработки дисбаланса классов и разработки гибридных подходов, которые объединяют сильные стороны различных методологий.

## Этапы метода

В разработанном методе рассматриваются две задачи анализа уязвимостей программного обеспечения: 1) обнаружение уязвимости (обнаружение потенциальных уязвимостей в образцах кода без указания их типов); 2) идентификация уязвимости (имеет ли предоставленный образец кода определенный тип уязвимости). Эти задачи охватывают различные уровни анализа уязвимости, представляя собой ключевые шаги в защите от уязвимостей программного обеспечения. Следовательно, анализ кода должен гарантированно определять уязвимости и их типы.

Задача обнаружения уязвимостей формулируется как бинарная классификация. Пусть имеется множество фрагментов программного кода  $C$ . Целью является создание детектора  $M: C \rightarrow V$  с выделением подмножества потенциально уязвимых фрагментов  $V = \{0, 1\}$ , где 1 обозначает уязвимый сегмент кода, а 0 обозначает безопасный сегмент.

Задача идентификации типа уязвимости направлена на определение того, имеет ли данный образец кода особый тип уязвимости (например, CWE). Для этого в исходном коде  $P$  необхо-

димо провести предварительную обработку кода и его отнесение к кластеру уязвимого. Учитывая, что исходный код  $P$  состоит из  $n$  фрагментов:  $C = \{c_1, c_2, \dots, c_n\}$ , каждый фрагмент  $c_i$ , представленный как вектор (эмбендинг)  $x_i \in R$ , должен быть нормализован  $\|x_i\| = 1$ , для чего используется метрика сходства – косинусного, евклидового или Махаланобиса – в качестве метрики расстояния для других сегментов.

Таким образом, в основе предлагаемого разработанного метода лежит гипотеза о сходстве уязвимых фрагментов кода (рис. 1). Для ее проверки реализуется двухступенчатый подход, включающий кластерный анализ и использование БЯМ с контекстной адаптацией.

Целью кластеризации является определение  $n$  фрагментов исходного кода в кластеры путем минимизирования внутрикластерной дисперсии:

$$\arg \min \sum_{i=1}^m \sum_{c_j \in k_i} \|x_i - \mu_i\|^2, \quad (1)$$

где  $\mu_i$  – центроид кластера уязвимых фрагментов кода. Это позволяет группировать потенциально уязвимые участки, даже если они находятся в разных частях программы.

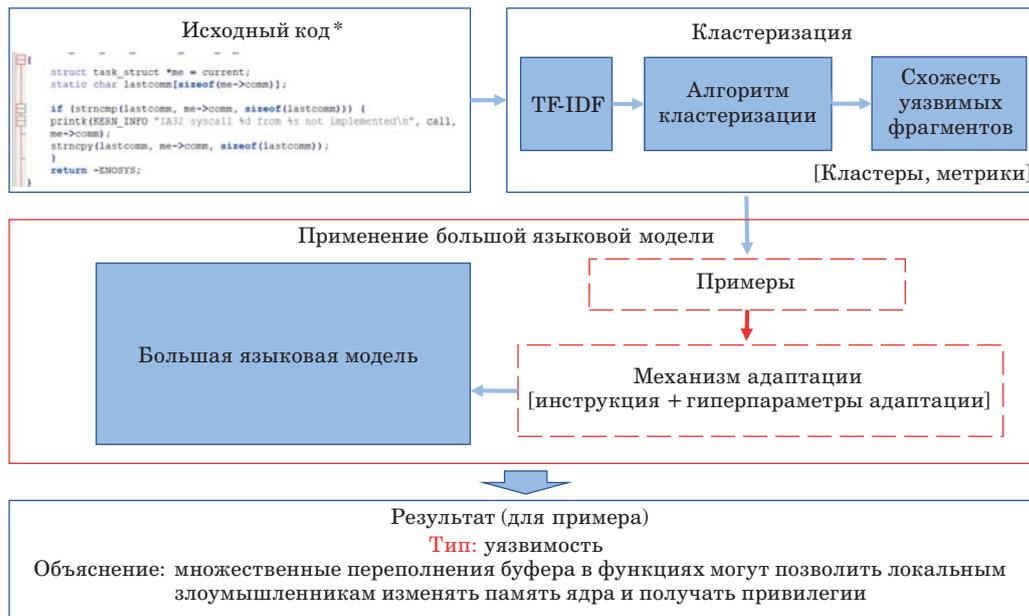
При определении принадлежности к кластеру каждому фрагменту кода присваивается значение метрики к ближайшему центроиду на основе наименьшего расстояния и обновляются центроиды соответствующим образом до тех пор, пока не произойдет схождение после  $T$  итераций.

Основная цель применения БЯМ для определения уязвимостей в программном коде – сформировать оптимальный набор центроидов кластеров уязвимых фрагментов кода и итеративно их корректировать для минимизации внутрикластерной дисперсии. Это связано со следующими особенностями выявления уязвимостей в программном коде:

1) хотя программа может включать в себя множество строк кода, только небольшая часть может быть подвержена уязвимостям. Следовательно, сосредоточение на этих критических фрагментах, составляющих семантику уязвимости, позволяет БЯМ эффективнее проводить анализ уязвимости;

2) большинство БЯМ основаны на трансформерах, которые, как известно, демонстрируют сниженную производительность при обработке/генерации обширного текста. Поэтому сокращение длины контекста позволяет улучшить сосредоточение на семантике уязвимости.

Использование БЯМ для увеличения признакового пространства уязвимости в программном коде заключается не только в использовании тестового образца программного кода, но и в фор-



\*Пример для CVE\_2004\_1151.

■ **Рис. 1.** Общая схема метода определения уязвимостей программного кода на основе кластерного анализа и контекстной адаптации БЯМ  
 ■ **Fig. 1.** General scheme of the method for determining vulnerabilities in program code based on cluster analysis and contextual adaptation of large language models

мировании набора из нескольких образцов уязвимостей кода, которые предоставляют примеры для БЯМ в целях выполнения контекстной адаптации обучения.

Каждый сегмент кода  $c_i$  анализируется в контексте окружающих утверждений  $T_i = \{t_{i-1}, t_i, t_{i+1}, \dots, t_{i+n}\}$ . Это могут быть логически связанные строки, функции, блоки кода, комментарии и др., а совместный вход в языковую модель задается как  $x_i = \text{concat}(t_{i-1}, t_i, t_{i+1}, \dots, t_{i+n})$ .

Языковая модель  $M$  преобразует вход в контекстно-зависимый эмбединг  $h_i = M(x_i)$ , а оценка вероятности уязвимости производится через функцию классификации:

$$p_i = \sigma(w^T h_i + b); \quad (2)$$

$$\hat{y}_i = \begin{cases} 1, & p_i \geq \tau \\ 0, & p_i < \tau \end{cases} \quad (3)$$

где  $p_i \in [0, 1]$ ;  $\sigma$  — сигмоида;  $w, b$  — параметры линейного слоя; порог  $\tau$  определяет, считать ли фрагмент кода уязвимым.

Для каждого фрагмента кода, попавшего в кластер, применяется БЯМ, которая анализирует не только сам фрагмент, но и его окружение — контекстуальные утверждения, влияющие на семантику:  $x_1 = (c_i, T_i)$ . Это позволяет уточнить вероятность наличия уязвимости, основываясь не только на структуре, но и на смыс-

ле кода. В частности, каждый образец состоит из инструкции, примера кода и шагов рассуждения. Итоговая функция предсказания объединяет структурные признаки из кластеров  $C$  и семантический контекст  $\hat{y}_i = F(V, h_i, x_i)$ .

Для примера с рис. 1 рассмотрим следующий фрагмент исходного кода ядра Linux:

```
struct task_struct *me = current;
static char lastcomm[sizeof(me->comm)];
if (strcmp(lastcomm, me->comm, sizeof(lastcomm))) {
    printk(KERN_INFO "IA32 syscall %d from %s not implemented\n", call, me->comm);
    strncpy(lastcomm, me->comm, sizeof(lastcomm));
}
return -ENOSYS;
```

Метод обрабатывает данный код поэтапно.

1. Предварительный анализ кода, при котором извлекаются имена переменных, вызовы функций и операторы. Фрагмент представляется в виде векторного пространства, которое далее поддается кластерному анализу. Кластеризация позволяет определить, принадлежит ли данный фрагмент к подмножеству ранее размеченных уязвимых фрагментов. Данный код входит в кластер, характеризующийся следующими признаками:

– использование фиксированных буферов (`char lastcomm[...]`);

– копирование данных из структуры (*me->comm*);

– повторное использование буфера без инициализации между итерациями.

Эти признаки указывают на структурное сходство с известными уязвимостями, связанными с переполнением буфера.

2. Большая языковая модель получает кодовый фрагмент с расширенным контекстом как в пределах функции, так и вне ее (включая структуры, переменные и комментарии близких в кластере примеров кода) и делает вывод, что потенциальная уязвимость связана с тем, что *me->comm* может быть не *null-terminated* и использование *strcpy* без явной защиты ведет к риску записи за границы буфера *lastcomm*.

### Экспериментальные исследования

Для проведения экспериментов было разработано программное обеспечение, реализующее предложенный метод обнаружения уязвимостей. Реализация выполнена на языке Python и развернута на выделенном сервере с поддержкой Docker-контейнеризации в целях обеспечения изоляции процессов и возможности масштабирования. Использовалась модульная архитектура, предусматривающая запуск БЯМ как микросервисов с REST API, что обеспечивает параллельную обработку нескольких запросов (обработку программных кодов) и многопользовательский доступ. Среднее время анализа одного проекта (до 150 строк кода) составляет от 12 до 15 с, включая этапы предобработки, кластерного анализа (кластеры в формате .pkl) и вывода БЯМ. Код решения и инструкция по воспроизведению экспериментов будут опубликованы после государственной регистрации интеллектуальной собственности.

*Исходные данные.* В исследовании на разных этапах использовались разные наборы данных. На этапе адаптации БЯМ применялся BigVul [36], а для проверки решений — CVEfixes [37].

BigVul — набор данных об уязвимостях кода C/C++ из проектов с открытым исходным кодом Github. Всего набор данных содержит 217 007 записей об изменениях кода (извлеченных из исправлений зафиксированных версий). Все записи разделены на уязвимые (если в функции были изменены строки с дефектами) и неуязвимые функции.

Второй набор данных CVEfixes создан на основе Национальной базы данных уязвимостей США. В первоначальном выпуске набор данных охватывает все опубликованные CVE по состоянию на 9 июня 2021 года. Набор данных организован как реляционная база и охватывает 5495 коммитов по исправлению уязвимостей в 1754 проектах с открытым исходным

кодом для всего 5365 CVE в 180 различных типах CWE. Набор данных включает исходный код до и после исправления 18 249 файлов и 50 322 функций.

*Предобработка исходного кода* включает в себя конвертирование в векторное пространство с помощью статистической меры TF-IDF [38], которая отражает значимость уязвимого сегмента кода  $v_i$  относительно коллекции программ  $C$ :

$$TF-IDF(v_j, c_i, C) = TF(v_j, c_i) \cdot IDF(v_j, C); \quad (4)$$

$$TF(v_j, c_i) = \frac{\text{Количество уязвимостей } v_j \text{ в сегменте кода } c_i}{\text{Общее количество сегментов в исходном коде}}; \quad (5)$$

$$IDF(v_j, C) = \log \left( \frac{\text{Общее количество исходных кодов в наборе}}{\text{Количество сегментов, содержащих уязвимость } v_j} \right). \quad (6)$$

*Алгоритм кластеризации данных, содержащих уязвимый код.* Существует множество алгоритмов кластеризации. Они различаются в первую очередь тем, как они измеряют «сходство» или «близость», и с какими типами признаков они работают. Алгоритм k-средних [39] понятен и прост в применении в контексте выявления признаков уязвимостей программного кода. Кластеризация k-средних создает кластеры, помещая центроиды внутрь пространства признаков. Каждая точка в наборе данных назначается кластеру того центроида, к которому она ближе всего. k — это количество центроидов (т. е. кластеров).

Для набора данных BigVul, используемого для валидации, количество кластеров соответствует категории CWE, а определение близости сегментов кода начинается с расчета метрики до центроида всех CWE. Пример расчета по метрике Махаланобиса представлен в табл. 1 для сегмента кода

```
def vulnerable_function(user_input):
    eval(user_input) # Потенциальная уязвимость инъекции кода
```

*Используемые большие языковые модели.* Табл. 2 содержит информацию о трех БЯМ, использованных в эксперименте.

*Инструкции для больших языковых моделей (Приложение)* включают три варианта: ролевой запрос, последовательность запросов с контекстной адаптацией БЯМ и запрос на основе кластерного анализа и контекстной адаптации БЯМ.

■ **Таблица 1.** Пример определения близости сегментов кода к центроидам уязвимых сегментов кода  
 ■ **Table 1.** Example of determining the proximity of code segments to the centroids of vulnerable code segments

№ п/п	ID CWE	Метрика схожести	№ п/п	ID CWE	Метрика схожести	№ п/п	ID CWE	Метрика схожести	№ п/п	ID CWE	Метрика схожести
1	119	0,0320	12	22	0,4360	23	369	0,2455	34	668	0,0000
2	125	0,0357	13	254	0,0487	24	399	0,0183	35	674	0,4075
3	134	0,4255	14	264	0,0354	25	400	0,2434	36	704	0,2191
4	16	0,0000	15	269	0,4345	26	404	0,4912	37	732	0,1006
5	17	0,0000	16	284	0,0999	27	415	0,1709	38	74	0,0000
6	172	0,0000	17	287	0,5000	28	416	0,0305	39	77	0,4112
7	189	0,0403	18	295	0,0000	29	476	0,0291	40	772	0,0996
8	19	0,1784	19	310	0,1067	30	502	0,0000	41	787	0,1013
9	190	0,0651	20	311	0,4628	31	59	0,2579	42	79	0,2453
10	20	0,0244	21	347	0,0000	32	611	0,4404	43	834	0,4382
11	200	0,0401	22	362	0,1044	33	617	0,2368	44	835	0,4531

■ **Таблица 2.** Используемые большие языковые модели  
 ■ **Table 2.** Large language models used

Название и версия	Организация и год анонсирования	Размер (количество параметров)	Размер контекста (количество токенов)
Meta-Llama-3-8B-Instruct	Meta, 18 апреля 2024 г.	8 млрд	128 000
Mistral-7B-Instruct-v0.3	Mistral AI SAS, 22 мая 2024 г.	7 млрд	8000–128 000
SOLAR-10.7B-Instruct-v1.0	Upstage AI, 13 декабря 2023 г.	10,7 млрд	128 000

Метрики оценивания определения уязвимостей в исходном коде. Для оценки качества обнаружения уязвимостей в исходном коде применяются классические метрики из области машинного обучения и анализа данных. Основными являются точность, полнота и F-мера, вычисляемые из матрицы несоответствия (confusion matrix).

Матрица несоответствия представляет собой таблицу, в которой фиксируются результаты сравнения предсказаний решения с истинными метками. Для задачи бинарной классификации (уязвимый/неуязвимый код) матрица включает следующие элементы:

	Фактически уязвимый	Фактически неуязвимый
Предсказано уязвимый	True Positive (TP)	False Positive (FP)
Предсказано неуязвимый	False Negative (FN)	True Negative (TN)

где:

– TP (истинно положительное): правильно определенный уязвимый участок кода;

- FP (ложно положительное): ошибочно помеченный безопасный участок как уязвимый;
- FN (ложно отрицательное): не обнаружена существующая уязвимость;
- TN (истинно отрицательное): корректно определено, что участок безопасен.

Точность показывает долю правильно предсказанных уязвимых участков среди всех участков, классифицированных как уязвимые:

$$\text{Точность} = \frac{TP}{TP + FP}. \quad (7)$$

Высокая точность характеризует число ложных срабатываний.

Полнота показывает, какую долю всех реально уязвимых участков модель смогла обнаружить:

$$\text{Полнота} = \frac{TP}{TP + FN}. \quad (8)$$

Высокая полнота означает, что модель способна выявлять большинство уязвимостей, даже если при этом возникают ложные срабатывания.

F-мера является гармоническим средним между точностью и полнотой, обеспечивая баланс между этими двумя характеристиками:

$$F_1 = 2 \cdot \frac{\text{Точность} \cdot \text{Полнота}}{\text{Точность} + \text{Полнота}} \quad (9)$$

F-мера особенно полезна в случаях, когда необходимо учитывать как количество обнаруженных уязвимостей, так и качество предсказаний.

Результаты экспериментов представлены в табл. 3 и на рис. 2.

■ **Таблица 3.** Результаты экспериментов  
 ■ **Table 3.** Experimental results

Модель	Режим	Точность	Полнота	F1-мера
Meta-Llama-3-8B-Instruct	РЗ	0,621	0,687	0,652
	ПЗ	0,710	0,769	0,738
	КлА+КА	0,758	0,805	0,781
Mistral-7B-Instruct-v0.3	РЗ	0,604	0,670	0,634
	ПЗ	0,692	0,740	0,712
	КлА+КА	0,749	0,802	0,776
SOLAR-10.7B-Instruct-v1.0	РЗ	0,640	0,698	0,667
	ПЗ	0,734	0,770	0,750
	КлА+КА	0,783	0,826	0,804

РЗ – ролевой запрос; ПЗ – последовательность запросов; (КлА+КА) – запрос на основе кластерного анализа и контекстной адаптации БЯМ.

### Обсуждение

Результаты экспериментов позволили сформулировать следующие закономерности:

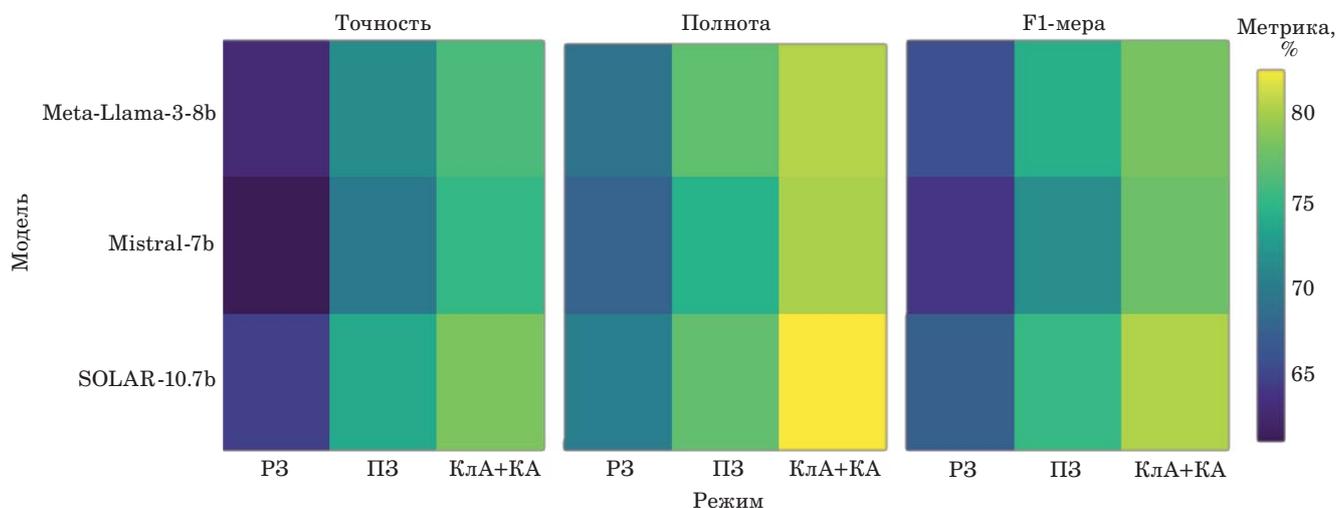
- все БЯМ способны выявлять уязвимости в коде;
- добавление примеров в запросы улучшает результаты классификации;
- использование наиболее похожих примеров при кластерном анализе и запрос на дополнительный анализ кода повышают эффективность моделей.

Результаты экспериментов доказывают, что комбинация различных техник формирования запросов может повысить качество обнаружения уязвимостей в исходном коде до 80 % и более.

Сравнительный анализ результатов экспериментов позволяет сделать следующие выводы:

- влияние механизмов формирования запросов и использование релевантных примеров улучшают способность моделей выявлять уязвимости, предоставляя контекст и примеры для обучения;
- размер модели влияет на результат, так как более крупные модели, такие как SOLAR-10.7B, обладают большей способностью к обобщению и выявлению сложных уязвимостей, однако и требуют больше вычислительных ресурсов;
- для достижения высокой точности в задачах обнаружения уязвимостей может потребоваться дополнительная настройка моделей на специализированных наборах данных, а также эффективных механизмов учета признаков в этих данных.

Для выделения вклада работы на фоне уже существующих решений был проведен сравнитель-



■ **Рис. 2.** Тепловая карта результатов экспериментов  
 ■ **Fig. 2.** Heat map of experimental results

ный анализ решений, таких как DeepWukong [40], VulBERTa [41] и GPT-анализаторы-ассистенты [42] (табл. 4).

Для детализации представим отличия предложенного подхода:

- от DeepWukong – вместо строгой привязки к графу потока управления применяется кластеризация кода по структурно-семантическому признаку, что снижает чувствительность к синтаксическим различиям и расширяет применимость к ранее неразмеченному коду, а использование БЯМ позволяет извлекать глобальный контекст, выходя за пределы локального анализа графа;

- от VulBERTa – не требует предварительной доразметки токенов и может применяться к неразмеченному коду, так как сначала происходит группировка схожих по структуре фрагментов, а затем контекстный анализ с помощью БЯМ;

- от GPT-ассистентов – предлагается интеграция контекста на уровне фрагмента кода, что позволяет учесть типовые зависимости (например, инициализацию переменных, вызов функций, структуру блоков).

Преимуществами предложенного метода являются отсутствие необходимости в токенизированной разметке, высокая устойчивость к синтаксическим вариациям и применимость к новому коду без дообучения. При этом указанные существующие решения служат важной основой для развития адаптивности и контекстной чувствительности определения уязвимостей в исходном коде.

■ **Таблица 4.** Сравнительный анализ аналогов

■ **Table 4.** Comparative analysis of analogues

Критерий	Решение			
	DeepWukong	VulBERTa	GPT-ассистент	Предложенный метод
Механизм	Графовая нейронная сеть	Трансформер-классификатор	Диалоговая БЯМ	Кластеризация + БЯМ
Учет структуры программы	+	–	–	+
Учет семантики	–	+	+	+
Требуется разметка	+	+	–	–
Поддержка контекста	–	–	Частично	+

Дополнительно следует отметить: нельзя с полной уверенностью утверждать, что части датасетов BigVul и CVEfixes не были применены для обучения используемых в эксперименте моделей. Это создает риск искусственного завышения показателей эффективности обнаружения уязвимостей в исходном коде.

Для устранения этих ограничений в дальнейшем планируется исследовать вопросы дообучения БЯМ на верифицированных, заведомо не пересекающихся подмножествах уязвимого и безопасного кода, или использовать темпорально отсеченные данные, обнаруженные после даты обучения модели, или проводить эксперименты на новых, ранее не встречавшихся в обучении уязвимостях из актуальных репозиториях кода.

### Заключение

Предложенный метод демонстрирует потенциал предварительной кластеризации фрагментов кода и применения больших языковых моделей в области автоматического обнаружения уязвимостей в исходном коде. Такое сочетание позволяет определять как явно выраженные, так и скрытые шаблоны уязвимостей, особенно в случаях, где локальные признаки недостаточны для точной классификации.

Результаты экспериментов на общедоступных датасетах BigVul и CVEfixes подтверждают эффективность предложенного решения. Так, достигнута точность до 78 % и полнота 82 %, что на 9–12 % превосходит показатели ряда существующих подходов. Преимуществом метода является устойчивость к синтаксическим вариациям кода, а также способность анализировать ранее неразмеченные фрагменты без необходимости ручной верификации.

В процессе исследования также определена значительная зависимость качества выявления уязвимостей от используемой стратегии формирования входных представлений и степени учета контекстной информации. Это подтверждает необходимость дальнейшей адаптации моделей с учетом особенностей языка программирования, а также перспективность применения мультязыковых решений для повышения универсальности метода.

Будущие направления исследований могут быть в области:

- дообучения БЯМ и использования дополненного поиска для выявления уязвимостей в исходном коде;
- развития специализированных токенизаторов, а также кастомизации эмбедингов;
- создания специфичных эвристик в фазе кластерного анализа;

- интеграции информации о CWE в запросы для возможного сочетания БЯМ с методами статического анализа;
- использования более крупных и разнообразных датасетов и меньших моделей;
- формирования более продвинутых техник для запросов, таких как самосогласованность (Self-Consistency) [43] и самопознание (Self-Discover) [44];
- оценивания безопасности и функциональности предложенных моделью исправлений кода.

## ПРИЛОЖЕНИЕ

### Инструкции для больших языковых моделей

#### Вариант 1. Ролевой запрос

Инструкция: Вы – эксперт по безопасности программного обеспечения. Вам будет предоставлен программный код. Если он содержит какие-либо уязвимости безопасности CWE, напишите Vulnerable. Если код не содержит никаких уязвимостей, напишите Not Vulnerable. Отформатируйте свой ответ как объект JSON с “label” в качестве ключа для статуса уязвимости и “cwe” в качестве номера найденной уязвимости.

Шаблон вывода: Исходный код: ```{code}```, вывод: ”label”.

#### Вариант 2. Последовательность запросов

Инструкция: Вы – эксперт по безопасности программного обеспечения. Вам будет предоставлен программный код. Если он содержит какие-либо уязвимости безопасности CWE, напишите Vulnerable. Если код не содержит никаких уязвимостей, напишите Not Vulnerable. Отформатируйте свой ответ как объект JSON

с “label” в качестве ключа для статуса уязвимости и “cwe” в качестве номера найденной уязвимости.

Продумайте ответ шаг за шагом и отвечайте только с помощью JSON.

Шаблон вывода:

Исходный код: ```{example\_0}```

Вывод: {label\_0}

Исходный код: ```{example\_1}```

Вывод: {label\_1}

Исходный код: ```{example\_2}```

Вывод: {label\_2}

Исходный код: ```{code}```

Конечный вывод: “label”.

### Вариант 3. На основе кластерного анализа и контекстной адаптации больших языковых моделей

Вы – эксперт по безопасности программного обеспечения. Вам будет предоставлен программный код. Если он содержит какие-либо уязвимости безопасности CWE, напишите Vulnerable. Если код не содержит никаких уязвимостей, напишите Not Vulnerable.

Проверьте, соответствует ли предоставленная метка (Vulnerable / Not Vulnerable) фактическому состоянию кода.

Если код уязвим:

- Определите конкретные уязвимости в коде.
- Объясните, почему эти уязвимости опасны.
- Опишите потенциальные последствия эксплуатации этих уязвимостей.

### Формат ответа:

– \*\*Метка\*\*: '{label}'.

– Если уязвим:

1. Опишите уязвимость как '{k\_mean\_example\_code}'.
2. Объясните, почему это опасно, как '{k\_mean\_example\_label}'.

## Литература

1. Jimmy F. N. U. Cyber security vulnerabilities and remediation through cloud security tools. *Journal of Artificial Intelligence General Science (JAIGS)*, 2024, vol. 2, no. 1, pp. 129–171. doi:10.60087/jaigs.v2i1.102
2. Fu M., Tantithamthavorn C., Le T., Kume Y., Nguyen V., Phung D., Grundy J. AIBugHunter: A practical tool for predicting, classifying and repairing software vulnerabilities. *Empirical Software Engineering*, 2024, vol. 29, no. 1, p. 4. doi:10.1007/s10664-023-10346-3
3. Jabbari M., Alibabaei A., Kavousi A., Rezvanjah M. Vulnerability analysis of the tanks of an oil refinery to fire-induced domino effects based on graph theory. *Iran Occupational Health*, 2021, vol. 18, no. 1, pp. 201–219. doi:10.52547/ioh.18.1.201

4. Shiri Harzevili N., Boaye Belle A., Wang J., Wang S., Jiang Z. M. (J.), Nagappan N. A systematic literature review on automated software vulnerability detection using machine learning. *ACM Computing Surveys*, 2024, vol. 57, no. 3, pp. 1–36. doi:10.1145/3699711
5. Prümmer J., van Steen T., van den Berg B. A systematic review of current cybersecurity training methods. *Computers & Security*, 2024, vol. 136, Article 103585. doi:10.1016/j.cose.2023.103585
6. Camacho N. G. The role of AI in cybersecurity: Addressing threats in the digital age. *Journal of Artificial Intelligence General Science (JAIGS)*, 2024, vol. 3, no. 1, pp. 143–154. doi:10.60087/jaigs.v3i1.75
7. Guo Y., Patsakis C., Hu Q., Tang Q., Casino F. Outside the comfort zone: Analysing LLM capabilities in software vulnerability detection. *European Symposi-*

- um on Research in Computer Security, 2024, pp. 271–289. doi:10.1007/978-3-031-70879-4\_14
8. Wang H., Tang Z., Tan S. H., Wang J., Liu Y., Fang H., Xia C., Wang Z. Combining structured static code information and dynamic symbolic traces for software vulnerability prediction. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.3639212
  9. Ferrag M. A., Battah A., Tihanyi N., Jain R., Maimut D., Alwahedi F., Lestable T., Thandi N. S., Mechri A., Debbah M., Cordeiro L. C. Secure falcon: Are we there yet in automated software vulnerability detection with LLMs? *IEEE Transactions on Software Engineering*, 2025, pp. 1248–1265. doi:10.1109/TSE.2025.3548168
  10. Zhao S., Yang Y., Wang Z., He Z., Qiu L. K., Qiu L. Retrieval augmented generation (rag) and beyond: A comprehensive survey on how to make your LLMs use external data more wisely. *arXiv preprint arXiv:2409.14924*, 2024. doi:10.48550/arXiv.2409.14924
  11. Patil R., Gudivada V. A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, 2024, vol. 14, no. 5, Article 2074. doi:10.3390/app14052074
  12. Chen L., Varoquaux G. What is the role of small models in the LLM era: A survey. *arXiv preprint arXiv:2409.06857*, 2024. doi:10.48550/arXiv.2409.06857
  13. Cao S., Sun X., Wu X., Lo D., Bo L., Li B., Liu W. Coca: Improving and explaining graph neural network-based vulnerability detection systems. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.3639168
  14. Chu Z., Wan Y., Li Q., Wu Y., Zhang H., Sui Y., Xu G., Jin H. Graph neural networks for vulnerability detection: A counterfactual explanation. *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis*, 2024, pp. 389–401. doi:10.1145/3650212.365213
  15. Senanayake J. M. D., Kalutarage H., Petrovski A., Piras L., Al-Kadri V. O. Defendroid: Real-time Android code vulnerability detection via blockchain federated neural network with XAI. *Journal of Information Security and Applications*, 2024, vol. 82, Article 103741. doi:10.1016/j.jisa.2024.103741
  16. Mahdavifar S., Saqib M., Fung B. C. M., Charland P., Walenstein A. VulEXplainer: XAI for vulnerability detection on assembly code. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2024, pp. 3–20. doi:10.1007/978-3-031-70378-2\_1
  17. Shareef F. RetailGPT: A fine-tuned LLM architecture for customer experience and sales optimization. *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, IEEE, 2024, pp. 1390–1394. doi:10.1109/ICSSAS64001.2024.10760685
  18. Berti A., Kourani H., Hafke H., Li C., Schuster D. Evaluating large language models in process mining: Capabilities, benchmarks, and evaluation strategies. *International Conference on Business Process Modeling, Development and Support*, 2024, pp. 13–21. doi:10.1007/978-3-031-61007-3\_2
  19. Ferrag M. A., Alwahedi F., Batah A., Cherif B. Generative ai and large language models for cyber security: All insights you need. *Available at SSRN 4853709*, 2024. doi:10.48550/arXiv.2405.12750
  20. Hassanin M., Moustafa N. A comprehensive overview of large language models (LLMs) for cyber defences: Opportunities and directions. *arXiv preprint arXiv:2405.14487*, 2024. doi:10.48550/arXiv.2405.14487
  21. Jelodar H., Bai S., Hamedi P., Mohammadian H., Razavi-Far R., Ghorbani A. Large Language Model (LLM) for software security: Code analysis, malware analysis, reverse engineering. *arXiv preprint arXiv:2504.07137*, 2025. doi:10.48550/arXiv.2504.07137
  22. Li Z., Dutta S., Naik M. LLM-assisted static analysis for detecting security vulnerabilities. *arXiv preprint arXiv:2405.17238*, 2024. doi:10.48550/arXiv.2405.17238
  23. Cao D., Jun W. LLM-CloudSec: Large Language Model empowered automatic and deep vulnerability analysis for intelligent clouds. *IEEE INFOCOM 2024-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2024, pp. 1–6. doi:10.1109/INFOCOMWKSHPS61880.2024.10620804
  24. Wu J., Yang S., Zhan R., Yuan Y., Wong D. F., Chao L. S. A survey on LLM-generated text detection: Necessity, methods, and future directions. *Computational Linguistics*, 2025, pp. 1–66. doi:10.1162/coli\_a\_00549
  25. Achiam J., Adler S., Agarwal S., Ahmad L., Akkaya I., Aleman F. L., Almeida D. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. doi:10.48550/arXiv.2303.08774
  26. Ersoy P., Erşahin M. Benchmarking Llama 3 70B for code generation: A comprehensive evaluation. *Orclever Proceedings of Research and Development*, 2024, vol. 4, no. 1, pp. 52–58.
  27. Song Y. Semantic attention and LLM-based layout guidance for text-to-image generation. *ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5. doi:10.1109/ICASSP49660.2025.10890155
  28. Manigrasso F., Schouten S., Morra L., Bloem P. Probing LLMs for logical reasoning. *International Conference on Neural-Symbolic Learning and Reasoning*, 2024, pp. 257–278. doi:10.1007/978-3-031-71167-1\_14
  29. Hashimoto W., Yasui M., Takeuchi K. Basic investigation of code edit distance measurement by CodeBERT. *Conference: 2023 15th International Congress*

- on *Advanced Applied Informatics Winter (II-AI-AAI-Winter)*, IEEE, 2023, pp. 13–18. doi:10.1109/IIAI-AAI-Winter61682.2023.00012
30. Zhang B., Zhang X., Zhang J., Yu J., Luo S., Tang J. CoT-based synthesizer: Enhancing LLM performance through answer synthesis. *arXiv preprint arXiv:2501.01668*, 2025. doi:10.48550/arXiv.2501.01668
31. Nam D., Macvean A., Hellendoorn V., Vasilescu B., Myers B. Using an llm to help with code understanding. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.363918
32. Wang P., Yuan N., Li Y. An integrated framework for data security using advanced machine learning classification and best practices. *Informatica*, 2025, vol. 49, no. 12, pp. 191–206. doi:10.31449/inf.v49i12.7838
33. Croft R., Babar M. A., Kholoosi M. M. Data quality for software vulnerability datasets. *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, IEEE, 2023, pp. 121–133. doi:10.1109/ICSE48619.2023.00022
34. Jain A., Patel H., Nagalapatti L., Gupta N., Mehta S., Guttula Sh., Mujumdar Sh., Afzal Sh., Mittal R. Sh., Munigala V. Overview and importance of data quality for machine learning tasks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3561–3562. doi:10.1145/3394486.3406477
35. Silva A., Saavedra N., Monperrus M. Gitbug-Java: A reproducible benchmark of recent Java bugs. *Proceedings of the 21st International Conference on Mining Software Repositories*, 2024, pp. 118–122. doi:10.1145/3643991.3644884
36. Fan J., Nguyen T. N. AC/C++ code vulnerability dataset with code changes and CVE summaries. *Proceedings of the 17th International Conference on Mining Software Repositories*, 2020, pp. 508–512. doi:10.1145/3379597.3387501
37. Bhandari G., Naseer A., Moonen L. CVEfixes: Automated collection of vulnerabilities and their fixes from open-source software. *Proceedings of the 17th International Conference on Predictive Models and Data Analytics in Software Engineering*, 2021, pp. 30–39. doi:10.1145/3475960.3475985
38. Sharma G., Singh P. Comparative study: Word2Vec Versus TF-IDF in software defect predictions. *International Conference on Data Science and Network Engineering*, Singapore, 2024, pp. 95–107. doi:10.1007/978-981-97-8336-6\_8
39. Yadav J., Sharma M. A review of K-mean algorithm. *Int. Journal of Engineering Trends and Technology*, 2013, vol. 4, no. 7, pp. 2972–2976. doi:10.1016/j.ins.2022.11.139
40. Cheng X., Wang H., Hua J., Xu G. Deepwukong: Statically detecting software vulnerabilities using deep graph neural network. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2021, vol. 30, no. 3, pp. 1–33. doi:10.1145/3436877
41. Hanif H., Maffei S. Vulberta: Simplified source code pre-training for vulnerability detection. *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2022, pp. 1–8. doi:10.1109/IJCNN55064.2022.9892280
42. Sun Y., Wu D., Xue Y., Liu H., Wang H., Xu Z., Xie X., Liu Y. GPTScan: Detecting logic vulnerabilities in smart contracts by combining GPT with program analysis. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.3639117
43. Ahmed T., Devanbu P. Better patching using LLM prompting, via self-consistency. *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, IEEE, 2023, pp. 1742–1746. doi:10.1109/ASE56229.2023.00065
44. Zhou P., Pujara J., Ren X., Chen X., Cheng H.-T., Le Q. V., Chi E. H., Zhou D., Mishra S., Zheng H. S. Self-discover: Large language models self-compose reasoning structures. *Advances in Neural Information Processing Systems*, 2024, vol. 37, pp. 126032–126058.

UDC 004.89

doi:10.31799/1684-8853-2025-4-58-70

EDN: XJDOLW

### Method for identifying software code vulnerabilities based on cluster analysis and contextual adaptation of large language models

R. N. Bakeev<sup>a</sup>, Head of the Center, orcid.org/0009-0008-3149-3836V. N. Kuzmin<sup>b</sup>, Dr. Sc., Military, Leading Researcher, orcid.org/0000-0002-6411-4336, vka@mil.ruA. B. Menisov<sup>b</sup>, Dr. Sc., Tech., Senior Lecturer, orcid.org/0000-0002-9955-2694T. R. Sabirov<sup>b</sup>, PhD, Tech., Senior Lecturer, orcid.org/0000-0002-6807-2954<sup>a</sup>Fund for Advanced Research, 22, bld. 3, Berezhkovskaya Emb., 121059, Moscow, Russian Federation<sup>b</sup>A. F. Mozhaiskii Military Space Academy, 13, Zhdanovskaia Emb., 197198, Saint-Petersburg, Russian Federation

**Introduction:** Detecting vulnerabilities in source code remains one of the priority tasks in the field of cybersecurity. Classical code analysis methods often do not take into account the execution context and do not scale well with the growth of program volumes. In the context of complex architectures and incomplete data annotations, a context-adapted approach is required that can identify vulnerabilities

based on semantic and structural analysis. **Purpose:** To develop a method for detecting vulnerabilities in program code using cluster analysis and contextual adaptation of large language models capable of taking into account not only the syntax, but also the semantic structure of programs. **Methods:** The developed approach combines clustering of code segments with the use of pre-trained language models adapted to the program code. To improve the efficiency, vulnerability features are identified, including both the vulnerable fragment itself and its context – control structures, variables, function calls. We carry out training and evaluation on labeled open datasets using pre-trained large language models. **Results:** The method makes it possible to automatically group code fragments by structural similarity, after which semantic analysis is performed using large language models capable of recognizing vulnerability patterns. Experiments have shown that the inclusion of contextual information significantly increases the efficiency of identifying vulnerabilities in the source code. On the BigVul and CVEfixes datasets, the proposed method achieved accuracy of up to 78% and recall of 82%, which is 9–12% higher than existing solutions. The method demonstrates resistance to syntactic variations and can be used to analyze previously unlabeled code. **Practical relevance:** The method is applicable in automatic source code analysis systems and can significantly reduce the cost of manual auditing, especially when analyzing large code bases. It can also be used for educational and research purposes to analyze vulnerability patterns. **Discussion:** The results confirm the effectiveness of using large language models in software security analysis tasks. A promising direction is to extend the method to other programming languages, as well as to study hybrid approaches involving graph neural networks. The question of arguing the model's decisions and automatically explaining the reasons for classifying a segment as vulnerable remains open.

**Keywords** – information security, vulnerability detection, large language models, cluster analysis, machine learning.

**For citation:** Bakeev R. N., Kuzmin V. N., Menisov A. B., Sabirov T. R. Method for identifying software code vulnerabilities based on cluster analysis and contextual adaptation of large language models. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 58–70 (In Russian). doi:10.31799/1684-8853-2025-4-58-70, EDN: XJDOLW

## References

1. Jimmy F. N. U. Cyber security vulnerabilities and remediation through cloud security tools. *Journal of Artificial Intelligence General Science (JAIGS)*, 2024, vol. 2, no. 1, pp. 129–171. doi:10.60087/jaigs.v2i1.102
2. Fu M., Tantithamthavorn C., Le T., Kume Y., Nguyen V., Phung D., Grundy J. AIBugHunter: A practical tool for predicting, classifying and repairing software vulnerabilities. *Empirical Software Engineering*, 2024, vol. 29, no. 1, p. 4. doi:10.1007/s10664-023-10346-3
3. Jabbari M., Alibabaei A., Kavousi A., Rezvanjah M. Vulnerability analysis of the tanks of an oil refinery to fire-induced dom-ino effects based on graph theory. *Iran Occupational Health*, 2021, vol. 18, no. 1, pp. 201–219. doi:10.52547/loh.18.1.201
4. Shiri Harzevili N., Boaye Belle A., Wang J., Wang S., Jiang Z. M. (J.), Nagappan N. A systematic literature review on automated software vulnerability detection using machine learning. *ACM Computing Surveys*, 2024, vol. 57, no. 3, pp. 1–36. doi:10.1145/3699711
5. Prümmer J., van Steen T., van den Berg B. A systematic review of current cybersecurity training methods. *Computers & Security*, 2024, vol. 136, Article 103585. doi:10.1016/j.cose.2023.103585
6. Camacho N. G. The role of AI in cybersecurity: Addressing threats in the digital age. *Journal of Artificial Intelligence General Science (JAIGS)*, 2024, vol. 3, no. 1, pp. 143–154. doi:10.60087/jaigs.v3i1.75
7. Guo Y., Patsakis C., Hu Q., Tang Q., Casino F. Outside the comfort zone: Analysing LLM capabilities in software vulnerability detection. *European Symposium on Research in Computer Security*, 2024, pp. 271–289. doi:10.1007/978-3-031-70879-4\_14
8. Wang H., Tang Z., Tan S. H., Wang J., Liu Y., Fang H., Xia C., Wang Z. Combining structured static code information and dynamic symbolic traces for software vulnerability prediction. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.3639212
9. Ferrag M. A., Battah A., Tihanyi N., Jain R., Maimut D., Alwahedi F., Lestable T., Thandi N. S., Mechri A., Debbah M., Cordeiro L. C. Secure falcon: Are we there yet in automated software vulnerability detection with LLMs? *IEEE Transactions on Software Engineering*, 2025, pp. 1248–1265. doi:10.1109/TSE.2025.3548168
10. Zhao S., Yang Y., Wang Z., He Z., Qiu L. K., Qiu L. Retrieval augmented generation (rag) and beyond: A comprehensive survey on how to make your LLMs use external data more wisely. *arXiv preprint arXiv:2409.14924*, 2024. doi:10.48550/arXiv.2409.14924
11. Patil R., Gudivada V. A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, 2024, vol. 14, no. 5, Article 2074. doi:10.3390/app14052074
12. Chen L., Varoquaux G. What is the role of small models in the LLM era: A survey. *arXiv preprint arXiv:2409.06857*, 2024. doi:10.48550/arXiv.2409.06857
13. Cao S., Sun X., Wu X., Lo D., Bo L., Li B., Liu W. Coca: Improving and explaining graph neural network-based vulnerability detection systems. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.3639168
14. Chu Z., Wan Y., Li Q., Wu Y., Zhang H., Sui Y., Xu G., Jin H. Graph neural networks for vulnerability detection: A counterfactual explanation. *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis*, 2024, pp. 389–401. doi:10.1145/3650212.365213
15. Senanayake J. M. D., Kalutarage H., Petrovski A., Piras L., Al-Kadri V. O. Defendroid: Real-time Android code vulnerability detection via blockchain federated neural network with XAI. *Journal of Information Security and Applications*, 2024, vol. 82, Article 103741. doi:10.1016/j.jisa.2024.103741
16. MahdaviFar S., Saqib M., Fung B. C. M., Charland P., Walenstein A. VulEXplaineR: XAI for vulnerability detection on assembly code. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2024, pp. 3–20. doi:10.1007/978-3-031-70378-2\_1
17. Shareef F. RetailGPT: A fine-tuned LLM architecture for customer experience and sales optimization. *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, IEEE, 2024, pp. 1390–1394. doi:10.1109/ICSSAS64001.2024.10760685
18. Berti A., Kourani H., Hafke H., Li C., Schuster D. Evaluating large language models in process mining: Capabilities, benchmarks, and evaluation strategies. *International Conference on Business Process Modeling, Development and Support*, 2024, pp. 13–21. doi:10.1007/978-3-031-61007-3\_2
19. Ferrag M. A., Alwahedi F., Batah A., Cherif B. Generative ai and large language models for cyber security: All insights you need. *Available at SSRN 4853709*, 2024. doi:10.48550/arXiv.2405.12750
20. Hassanin M., Moustafa N. A comprehensive overview of large language models (LLMs) for cyber defences: Opportunities and directions. *arXiv preprint arXiv:2405.14487*, 2024. doi:10.48550/arXiv.2405.14487
21. Jelodar H., Bai S., Hamed P., Mohammadian H., Raza-vi-Far R., Ghorbani A. Large Language Model (LLM) for software security: Code analysis, malware analysis, reverse engineering. *arXiv preprint arXiv:2504.07137*, 2025. doi:10.48550/arXiv.2504.07137
22. Li Z., Dutta S., Naik M. LLM-assisted static analysis for detecting security vulnerabilities. *arXiv preprint arXiv:2405.17238*, 2024. doi:10.48550/arXiv.2405.17238
23. Cao D., Jun W. LLM-CloudSec: Large Language Model empowered automatic and deep vulnerability analysis for intelligent clouds. *IEEE INFOCOM 2024-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2024, pp. 1–6. doi:10.1109/INFOCOM-WKSHPS61880.2024.10620804
24. Wu J., Yang S., Zhan R., Yuan Y., Wong D. F., Chao L. S. A survey on LLM-generated text detection: Necessity, methods, and future directions. *Computational Linguistics*, 2025, pp. 1–66. doi:10.1162/coli\_a\_00549
25. Achiam J., Adler S., Agarwal S., Ahmad L., Akkaya I., Aleman F. L., Almeida D. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. doi:10.48550/arXiv.2303.08774

26. Ersoy P., Erşahin M. Benchmarking Llama 3 70B for code generation: A comprehensive evaluation. *Orclever Proceedings of Research and Development*, 2024, vol. 4, no. 1, pp. 52–58.
27. Song Y. Semantic attention and LLM-based layout guidance for text-to-image generation. *ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5. doi:10.1109/ICASSP49660.2025.10890155
28. Manigrasso F., Schouten S., Morra L., Bloem P. Probing LLMs for logical reasoning. *International Conference on Neural-Symbolic Learning and Reasoning*, 2024, pp. 257–278. doi:10.1007/978-3-031-71167-1\_14
29. Hashimoto W., Yasui M., Takeuchi K. Basic investigation of code edit distance measurement by CodeBERT. *Conference: 2023 15th International Congress on Advanced Applied Informatics Winter (IIAI-AAI-Winter)*, IEEE, 2023, pp. 13–18. doi:10.1109/IIAI-AAI-Winter61682.2023.00012
30. Zhang B., Zhang X., Zhang J., Yu J., Luo S., Tang J. CoT-based synthesizer: Enhancing LLM performance through answer synthesis. *arXiv preprint arXiv:2501.01668*, 2025. doi:10.48550/arXiv.2501.01668
31. Nam D., Macvean A., Hellendoorn V., Vasilescu B., Myers B. Using an llm to help with code understanding. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.363918
32. Wang P., Yuan N., Li Y. An integrated framework for data security using advanced machine learning classification and best practices. *Informatica*, 2025, vol. 49, no. 12, pp. 191–206. doi:10.31449/inf.v49i12.7838
33. Croft R., Babar M. A., Kholoosi M. M. Data quality for software vulnerability datasets. *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, IEEE, 2023, pp. 121–133. doi:10.1109/ICSE48619.2023.00022
34. Jain A., Patel H., Nagalapatti L., Gupta N., Mehta S., Guttula Sh., Mujumdar Sh., Afzal Sh., Mittal R. Sh., Munigala V. Overview and importance of data quality for machine learning tasks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3561–3562. doi:10.1145/3394486.3406477
35. Silva A., Saavedra N., Monperrus M. Gitbug-Java: A reproducible benchmark of recent Java bugs. *Proceedings of the 21st International Conference on Mining Software Repositories*, 2024, pp. 118–122. doi:10.1145/3643991.3644884
36. Fan J., Nguyen T. N. AC/C++ code vulnerability dataset with code changes and CVE summaries. *Proceedings of the 17th International Conference on Mining Software Repositories*, 2020, pp. 508–512. doi:10.1145/3379597.3387501
37. Bhandari G., Naseer A., Moonen L. CVEfixes: Automated collection of vulnerabilities and their fixes from open-source software. *Proceedings of the 17th International Conference on Predictive Models and Data Analytics in Software Engineering*, 2021, pp. 30–39. doi:10.1145/3475960.3475985
38. Sharma G., Singh P. Comparative study: Word2Vec Versus TF-IDF in software defect predictions. *International Conference on Data Science and Network Engineering*, Singapore, 2024, pp. 95–107. doi:10.1007/978-981-97-8336-6\_8
39. Yadav J., Sharma M. A review of K-mean algorithm. *Int. Journal of Engineering Trends and Technology*, 2013, vol. 4, no. 7, pp. 2972–2976. doi:10.1016/j.ins.2022.11.139
40. Cheng X., Wang H., Hua J., Xu G. Deepwukong: Statically detecting software vulnerabilities using deep graph neural network. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2021, vol. 30, no. 3, pp. 1–33. doi:10.1145/3436877
41. Hanif H., Maffei S. Vulberta: Simplified source code pre-training for vulnerability detection. *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2022, pp. 1–8. doi:10.1109/IJCNN55064.2022.9892280
42. Sun Y., Wu D., Xue Y., Liu H., Wang H., Xu Z., Xie X., Liu Y. GPTScan: Detecting logic vulnerabilities in smart contracts by combining GPT with program analysis. *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13. doi:10.1145/3597503.3639117
43. Ahmed T., Devanbu P. Better patching using LLM prompting, via self-consistency. *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, IEEE, 2023, pp. 1742–1746. doi:10.1109/ASE56229.2023.00065
44. Zhou P., Pujara J., Ren X., Chen X., Cheng H.-T., Le Q. V., Chi E. H., Zhou D., Mishra S., Zheng H. S. Self-discover: Large language models self-compose reasoning structures. *Advances in Neural Information Processing Systems*, 2024, vol. 37, pp. 126032–126058.

**АРХИПЦЕВ**  
**Евгений**  
**Дмитриевич**



Аспирант кафедры информационных систем Санкт-Петербургского государственного электротехнического университета «ЛЭТИ». В 2022 году окончил магистратуру Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» по направлению «Информационные системы и технологии». Является автором десяти научных публикаций. Область научных интересов – распределенные вычислительные системы. Эл. адрес: lokargenia@gmail.com

**БАКЕЕВ**  
**Руслан**  
**Надирович**



Руководитель Центра искусственного интеллекта Фонда перспективных исследований, Москва. В 2013 году окончил Московский государственный университет им. М. В. Ломоносова по специальности «Прикладная математика и информатика», в 2016 году – Финансовый университет при Правительстве Российской Федерации по специальности «Менеджмент». Область научных интересов – методы машинного обучения, оптимальное управление, вычислительные методы. Эл. адрес: bakeev.ruslan@yandex.ru

**ВОЛОДАРЕЦ**  
**Никита**  
**Витальевич**



Доцент, проректор по научно-педагогической работе Донецкой академии транспорта. В 2009 году окончил Донецкий институт железнодорожного транспорта по специальности «Подвижной состав и специальная техника железнодорожного транспорта». В 2016 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 250 научных публикаций, в том числе четырех монографий. Область научных интересов – интеллектуальные системы управления, автономные транспортные системы, глубокое обучение и нейронные сети в компьютерном зрении, оптимизация нейросетей для embedded-систем и др. Эл. адрес: volodarets.nikita@yandex.ru

**ДУМАЕВ**  
**Ринат**  
**Ильгизович**



Аспирант Высшей школы программной инженерии Санкт-Петербургского политехнического университета Петра Великого. В 2023 году окончил магистратуру Санкт-Петербургского политехнического университета Петра Великого по специальности «Программная инженерия». Является автором семи научных публикаций. Область научных интересов – компьютерное зрение, глубокое обучение, искусственный интеллект. Эл. адрес: dumayevrinat@gmail.com

**ЖУКОВ**  
**Станислав**  
**Владимирович**



Аспирант кафедры математического моделирования и информационных технологий Тамбовского государственного университета им. Г. Р. Державина. В 2021 году окончил Тамбовский государственный университет им. Г. Р. Державина по направлению «Прикладная информатика». Является автором 11 научных публикаций и шести свидетельств о регистрации программы для ЭВМ. Область научных интересов – информационные технологии, веб-разработка, математическое и компьютерное моделирование динамических систем. Эл. адрес: i@coder-stas.ru

**КОВАЛЕВ**  
**Сергей**  
**Владимирович**



Профессор кафедры математического моделирования и информационных технологий Тамбовского государственного университета им. Г. Р. Державина. В 2006 году окончил Тамбовский государственный технический университет по специальности «Технологические машины и оборудование», в 2014 году – по специальности «Прикладная математика и информатика». В 2015 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 350 научных публикаций, в том числе патентов на изобретения и свидетельств о государственной регистрации программы для ЭВМ. Область научных интересов – математическое и компьютерное моделирование процессов и систем, многокритериальная оптимизация. Эл. адрес: sseedd@mail.ru

**КОВАЛЕВА**  
**Ольга**  
**Александровна**



Профессор кафедры математического моделирования и информационных технологий Тамбовского государственного университета им. Г. Р. Державина. В 2005 году окончила Тамбовский государственный университет им. Г. Р. Державина по специальности «Математика». В 2019 году защитила диссертацию на соискание ученой степени доктора технических наук. Является автором более 250 научных публикаций и более 40 объектов интеллектуальной собственности. Область научных интересов — математическое и компьютерное моделирование динамических систем. Эл. адрес: solomina-oa@yandex.ru

**ЛЕВАШОВА**  
**Татьяна**  
**Викторовна**



Старший научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского Федерального исследовательского центра РАН. В 1986 году окончила Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Электронные вычислительные машины». В 2009 году защитила диссертацию на соискание ученой степени кандидата технических наук. Является автором более 200 научных публикаций. Область научных интересов — управление знаниями, инженерия онтологий, управление онтологиями, управление контекстом, социокриберфизические системы. Эл. адрес: tatiana.levashova@iias.spb.ru

**МОЛОДЯКОВ**  
**Сергей**  
**Александрович**



Профессор Высшей школы программной инженерии Санкт-Петербургского политехнического университета Петра Великого. В 1981 году окончил Ленинградский политехнический институт им. М. И. Калинина по специальности «Автоматизированные системы управления». В 2012 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 250 научных публикаций, в том числе семи монографий. Область научных интересов — методы проектирования гибридных процессоров, обработка сигналов, распознавание изображений и обработка видеoinформации. Эл. адрес: samolodyakov@mail.ru

**КУЗЬМИН**  
**Владимир**  
**Никифорович**



Профессор, ведущий сотрудник управления Военно-исследовательского (научного) института Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург. В 1976 году окончил Военный инженерный Краснознаменный институт им. А. Ф. Можайского по специальности «Автоматизированные системы управления». В 1996 году защитил диссертацию на соискание ученой степени доктора военных наук. Является автором 210 научных публикаций. Область научных интересов — теория и практика внедрения технологий ИИ, методы и стратегии повышения безопасности больших организационно-технических систем в условиях воздействия деструктивных факторов различной природы. Эл. адрес: vka@mail.ru

**МЕНИСОВ**  
**Артем**  
**Бакытжанович**



Старший преподаватель кафедры систем сбора и обработки информации Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург. В 2013 году окончил Военно-космическую академию им. А. Ф. Можайского по специальности «Информационные системы и технологии». В 2024 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором более 100 научных публикаций. Область научных интересов — искусственный интеллект, машинное обучение, нейронные сети, информационная безопасность. Эл. адрес: men.arty@yandex.ru

**САБИРОВ**  
**Тимур**  
**Римович**



Старший преподаватель кафедры систем сбора и обработки информации Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург. В 2009 году с отличием окончил Военно-космическую академию им. А. Ф. Можайского по специальности «Программное обеспечение вычислительной техники и автоматизированных систем». В 2017 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 75 научных публикаций. Область научных интересов — искусственный интеллект, интеллектуальные системы поддержки принятия решений. Эл. адрес: rowing-team@mail.ru

**САЦЮК**  
**Александр**  
**Владимирович**



Доцент кафедры автоматки, телемеханики и вычислительной техники, руководитель лаборатории искусственного интеллекта Донецкого института железнодорожного транспорта. В 2009 году окончил Донецкий институт железнодорожного транспорта по специальности «Автоматика и автоматизация на транспорте». В 2019 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 43 научных публикаций и трех патентов на изобретения. Область научных интересов — интеллектуальные системы управления, автономные транспортные системы, глубокое обучение и нейронные сети в компьютерном зрении и др.  
Эл. адрес: alexandrsatsuk@gmail.com

**ТАТАРНИКОВА**  
**Татьяна**  
**Михайловна**



Профессор, директор Института информационных технологий и программирования Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 1993 году окончила Восточно-Сибирский технологический институт по специальности «Электронно-вычислительные машины, комплексы, системы и сети». В 2007 году защитила диссертацию на соискание ученой степени доктора технических наук. Является автором более 300 научных публикаций. Область научных интересов — инфокоммуникации, взаимодействие неоднородных сетей, анализ данных.  
Эл. адрес: tm-tatarn@yandex.ru

**УТКИН**  
**Лев**  
**Владимирович**



Профессор, главный научный сотрудник Высшей школы технологий искусственного интеллекта Санкт-Петербургского политехнического университета Петра Великого. В 1986 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматика и телемеханика». В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 300 научных публикаций. Область научных интересов — машинное обучение, описание неопределенности, принятие решений, теория надежности.  
Эл. адрес: utkin\_lv@spbstu.ru

**СМИРНОВ**  
**Александр**  
**Викторович**



Профессор, главный научный сотрудник, заведующий лабораторией интегрированных систем автоматизации Санкт-Петербургского Федерального исследовательского центра РАН. В 1979 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Системы автоматического управления». В 1994 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 350 научных публикаций. Область научных интересов — концептуальное моделирование, управление знаниями, управление контекстом, конфигурирование систем, системы поддержки принятия решений, социкиберфизические системы.  
Эл. адрес: smir@iias.spb.su

**ТЕСЛЯ**  
**Николай**  
**Николаевич**



Старший научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского Федерального исследовательского центра РАН. В 2013 году окончил Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» по специальности «Компьютерная безопасность». В 2015 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором более 80 научных публикаций. Область научных интересов — управление знаниями, инженерия онтологий, слияние онтологий, социкиберфизические системы, поддержка принятия решений, практическое применение кооперативных игр, геoinформационные системы.  
Эл. адрес: teslya@iias.spb.su

## Уважаемые авторы!

### При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Текст рукописи должен быть оригинальным, а цитирование и самоцитирование корректно оформлено.

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, ORCID и электронный адрес одного из авторов. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы. Предоставляйте подрисуночные подписи и названия таблиц на русском и английском языках.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени.

Простые **формулы** набирайте в Word, сложные с помощью редактора Mathtype или Equation. Для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта в Mathtype никогда не пользуйтесь вкладкой Other, Smaller, Larger, используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; пробелы в формуле ставьте только после запятой при перечислении с помощью Ctrl+Shift+Space (пробел); не отделяйте пробелами знаки: + = - ×, а также пространство внутри скобок; для выделения греческих символов в Mathtype полужирным начертанием используйте Style → Other → bold.

Для набора формул в Word никогда не используйте вкладки: «Уравнение», «Конструктор», «Формула» (на верхней панели: «Вставка» — «Уравнение»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Подробнее см. <http://i-us.ru/index.php/ius/author-guide>

### Иллюстрации:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (\*.vsd, \*.vsdx); Adobe Illustrator (\*.ai); Coreldraw (\*.cdr, версия не выше 15); Excel (\*.xls); Word (\*.docx); AutoCad, Matlab (экспорт в PDF, EPS, SVG, WMF, EMF); Компас (экспорт в PDF); веб-портал DRAW.IO (экспорт в PDF); Inkscape (экспорт в PDF);

— фото и растровые — в формате \*.tif, \*.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисуночных подписей и названий таблиц на русском и английском языках обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

### В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате \*.tif, \*.png, \*.jpg с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение;

— экспортное заключение.

**Список литературы** составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц, doi;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц, doi;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (<http://i-us.ru/paperrules>): Литература и References.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Руководство для авторов» — <http://i-us.ru/index.php/ius/author-guide>.

## Контакты

Куда: 190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ

Кому: Редакция журнала «Информационно-управляющие системы»

Тел.: (812) 494-70-02

Эл. почта: [ius.spb@gmail.com](mailto:ius.spb@gmail.com)

Сайт: [www.i-us.ru](http://www.i-us.ru)