UDC 004.272
DOI: 10.20535/SRIT.2308-8893.2024.2.06

AN IMPROVED APPROACH TO ORGANISING MOBILE EDGE COMPUTING IN A 5G NETWORK

A. ASTRAKHANTSEV, L. GLOBA, O. FEDOROV, D. DEGTIAROV, Y. ROMANKO, K. ROMANII

Abstract. Mobile edge computing is an important element in ensuring the efficiency of the 5G network as a whole, as it enables data storage and computing at the network edge. Existing solutions do not fully address the issues of load distribution between computing nodes, and most solutions do not offer methods for verifying computations and controlling errors. Accordingly, this paper aims to develop an approach to the organization of mobile edge computing in a 5G mobile network that would authenticate distribution servers and computing nodes, manage the process of distributing computing nodes, have a procedure for verifying the correctness of calculations, and take into account the parameters of computing nodes during distribution. To achieve this goal, we propose to use the developed method. The method of load balancing and selection of computing nodes for edge computing via 5G allows for identifying available nodes and distributing computing blocks among them. It also provides mutual authentication of elements and includes a method of data verification and error detection for the MEC system. The provided solution allows for controlling errors during calculations and protecting the server from incorrect data. These methods are optimized according to minimum network resources and computing time criteria. These improvements increase the efficiency of mobile edge computing in a 5G network.

Keywords: 5G network, mobile edge computing, task allocation scheme, call flow, load balancing, task verification.

INTRODUCTION

The new 5G cellular networks are expected to face a sharp increase in mobile traffic and IoT user demands due to the massive growth in the number of mobile devices and the emergence of new computing applications. Running resource-intensive computing applications on resource-constrained mobile devices has recently become a major challenge, given the stringent requirements for computing time and the limited storage capacity of the devices.

Cloud computing allows you to store and process data on remote servers. A large number of different applications that generate an ever-increasing amount of data, which significantly increases network latency, uses them and places differentiated demands on data security and manageability. Mobile Edge Computing (MEC) technology can help prevent these problems from getting worse.

Mobile edge computing has recently emerged as a key technology to overcome these challenges, as it enables the provision of cloud computing services such as data storage and computing at the edge of the network. MECs have the potential to run computationally intensive applications such as augmented and virtual reality [1]. MEC is also an important component of the Internet of Things (IoT), as it allows to reduce the power consumption of mobile devices.

Mobile edge computing is a data management technology that involves storing and processing data close to the source. This allows for faster response to real-time computing needs and helps to guarantee the availability of information. In general, MEC is a decentralized computing infrastructure in which some signal processing, storage, management and computing applications are distributed in the most efficient and logical way between the data source and the cloud [2]. Mobile edge computing extends the concept of cloud computing by bringing the benefits of the cloud closer to users in the form of the network edge, which provides lower end-to-end latency.

The goal of the presented work is to organize mobile edge computing in a 5G network by performing authentication of distribution servers and computing nodes. It is also necessary to ensure the management of the process of distributing computing units, including the procedure for checking the correctness of calculations and taking into account the parameters of computing nodes during distribution.

In this regard, the following tasks were solved within the framework of an improved approach to the organization of mobile edge computing in the 5G network:

- Development of a method for load balancing and selection of computing nodes for MEC. The implementation of this method should not require additional physical elements in the network.
- Developing a method for data verification and error detection, as a computing node may report incorrect calculation results.
- Determining a method of mutual authentication for different types of equipment in the 5G network for the process of mobile edge computing without the use of a trusted third party.

At the same time, there are currently no existing technical solutions that would solve all of the above problems.

ANALYSIS OF EXISTING SOLUTIONS FOR MEC IN THE 5G NETWORK

The problems that arise when organizing mobile edge computing covered in a large number of publications. For example, [3] describes a typical MEC architecture and its main elements, as well as the problems associated with the distribution of computing tasks. Paper [4] focuses on the problems of transmission delay and computation delay with a large number of IoT devices. It also analyses the possibility of overloading peripheral clouds due to the spatially heterogeneous distribution of IoT tasks. To address these issues, it use game-theoretic methods to investigate load balancing problems to minimize transmission and computation delays in the task distribution process, given the limited bandwidth and computing resources in the edge clouds.

Work [5] solves a more complex problem of parallel offloading and load balancing with several shared MEC servers and delay-sensitive load. A similar problem is solved in [6], but it proposes a two-level model of task distribution with delay minimization and computational cost estimation. In [5], a long-term stochastic programming problem with an average system cost is formulated under the conditions of stability of the battery level and delay constraints.

Another work by the same authors [7] partially solves one of the problems studied in proposed research — it helps to create a secure reward mechanism using blockchain technology that can help to balance the load between computing nodes.

Some works [8] propose to solve the problem of clustering and load balancing based on the charge level of computing nodes and geolocation tags.

In addition to the solutions that solve the problem of load distribution between the computing nodes of the MEC and which are presented in the above publications, it is necessary analyzing patented solutions separately. For example, patent [9] proposes a cloud platform with a pool of resources that connects to the main network via a transmission network. This solution requires a special distribution hub (RRH). In this article, to eliminate this drawback, it is proposed to use a flat structure divided into zones and use the base station as an arbiter (no additional equipment is required).

Patent [10] proposes a solution based on the availability of a map that stores information about the location, computing power and available storage of each of the computing nodes. The disadvantage of the solution is the need for a dynamically updated map with a list of nodes. This disadvantage can be overcome by using a broadcast of the request from the MEC server. In this case, the response of the base station to the computing nodes allows not to use the map, route table or database — saving network resources.

The patent [11] is devoted to determining the optimal number of required physical resource blocks during distribution, while the procedure for selecting computing nodes is not described. In [12], the distribution of computing tasks and resources is based on reducing the failure rate during handover, but the procedure for allocating network resources is also not described and there is no data verification and error control. The solution [13] offers a centralized implementation of edge computing, where a central computer or a cloud macro base station will perform the main distribution tasks. This requires additional costs. In addition, this solution uses only delay as a distribution criterion. The patent [14] also requires a hierarchical structure and does not provide for the identification and authentication of computing nodes. In addition, this solution lacks data verification and error control.

The patent [15] describes only the process of creating a session for mobile edge computing, does not provide solution for data verification and error control, and does not describe procedures for load balancing and selection of computing nodes.

Solutions [16; 17] do not provide for security measures (no identification and authentication). In addition, in both cases, a central database is required. In [18], it is described how a computing node should be rewarded for a completed operation, but the data verification procedure is not described and there is no support for the 5G network, as well as no secure channel for transferring rewards.

To summarize, most solutions for selecting computing nodes and load balancing use additional physical elements, which requires additional costs. The considered technical solutions require dynamically updated maps or databases, which requires additional network resources and increases the load on the network. In addition, many of these publications lack authentication procedures for participants.

PROBLEM STATEMENT

Let us identify the main participants in the process of distributed mobile edge computing (Table) and their functions according to the approach proposed in this paper. A similar list of process participants, but with a different set of functions, is given in [3; 19; 20].

Main participants in the process of distributed edge computing

Participant marking	Participant functions and components
 MEC Server	MEC Server: gather data flow from one/multiple sensors; has 5G supported radio module; run MEC supported application; has identity and billing entity.
 Computing Node	Computing Node: process MEC Server Application Programming Interface (API) call; has 5G supported radio module; has CPU that support operability of MEC framework; has identity and billing entity.
 Base station (Cell)	Cell: assign radio resource, verify identity, sign transaction, secure connection; support computing unit selection; support peer-to-peer communication; secure and sign transaction MEC Server → Computing node;

The solution of the tasks set in this paper done by simulation and mathematical modelling for the architecture shown in Table, taking into account the shortcomings of existing solutions discussed in the previous section. In order to prepare the proposed technical solution, a set of input data, a set of constraints, dependencies between them, and a set of output values were determined. Let us consider them in more detail.

Let the following data received at the input of the load balancing system for distributed boundary computing:

n — a set of computing nodes available to the MEC with computing capacities r_i and an initial level of trust d_i;

d_i — initial level of trust in the computing node;

$p(x_i)$ — the probability of an error during calculations by the i-th node;

T_p — time for the distribution of computing tasks;

Res — the amount of network resources involved in the distribution of tasks;

T_{o} — the expected calculation time;

T_{def} — the restriction for the expected calculation time;

V — the amount of calculations to be performed.

When distributing computational tasks for execution, it is necessary to ensure the minimum probability of calculation error $p(n)$ and minimize the network resources used during their distribution and processing in computing nodes: $(\text{Res} \to \min)$, subject to restrictions on the expected computing time $(T_o \le T_{\text{def}})$. A weighted average used to determine the probability of a calculation error:

$$p(n) = \frac{1}{n}\sum_{i}^{n} p(x_i).$$

As a result, the proposed method should provide:

$y \in n(0,n)$ — a set of devices that perform the calculation of distributed computing with load balancing based on capacity r_i;

$w \in n(0, n-y)$ — a setoff additional devices that will provide redundancy and reliability of distributed computing;

Δd_i — change in the level of trust in the *i*-th node based on the results of its work.

The expected calculation time will consist directly of the calculation time and the time for distributing the calculation tasks defined as:

$$T_o = \frac{V}{nr} + T_p,$$

or considering the set of devices that perform the calculation and the set of additional devices:

$$T_o = \frac{V}{\sum_i^y (y_i r_i) + \sum_i^w (w_i r_i)} + T_p.$$

In this paper, we propose a method for organizing distributed computing that performs the following steps:

- a broadcast request from the MEC server to perform distributed computing;
- a response from at least one computing node to the base station containing a set of parameters (request ID, timestamp, etc.);
- the base station checks the available resources and provides network parameters for the MEC session of at least one computing node, which will allow for further point-to-point connection;
- the MEC server can verify the results of the calculations by means of data validation, mirroring and control code.

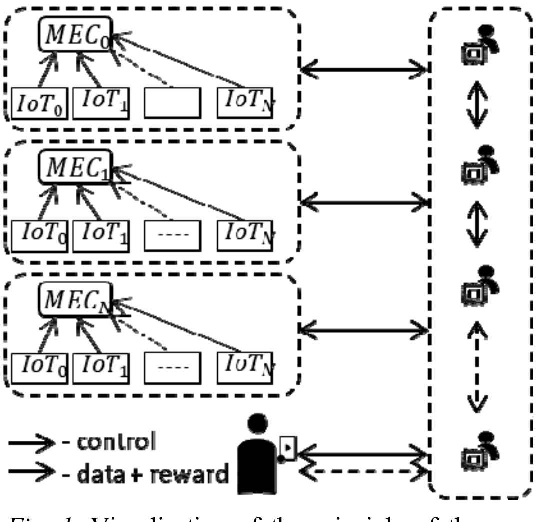

Fig. 1. Visualization of the principle of the proposed method

The essence of the method described above showed in Figs. 1, 2.

The proposed approach is based on two new methods:

- a method of load balancing and selection of computing nodes for edge computing via 5G, which allows to identify available nodes and distribute computing blocks among them, as well as provide mutual authentication of elements;
- a method of data verification and error detection for the MEC system, which allows to control the occurrence of errors during calculations, protect the server from incorrect data, and prioritize and reward nodes based on the results of the calculations performed.

Let us consider the principle of the proposed methods in more detail.

METHOD OF LOAD BALANCING AND SELECTION OF COMPUTING NODES IN A 5G NETWORK

The principle of the load balancing method includes two stages: 1) the authentication stage and 2) the point-to-point channel establishment and the computation and verification stage.

The first stage of authentication and channel creation (Fig. 2) involves the following steps:

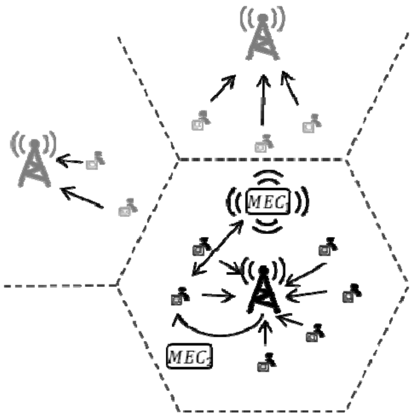

Fig. 2. The sequence of actions at the first stage of distributed computing — authentication and channel creation

1. The MEC server broadcasts a request to perform calculations with the following information:

- MEC identifier (temporary or permanent identifier);
- type of calculation.

2. Each computing node upon receiving paging device, reply to base station (cell) with:

$$E = F(C_{id}, T_{last}, E_{id}),$$

where (C_{id}) — identifier of the serving cell (base station); (T_{last}) — the timestamp of the last received slot for performing calculations; (E_{id}) — network identifier (temporary or permanent).

3. The base station selects computing nodes and assigns a radio channel:

- selects a computing node based on the received values of E;
- assigns a radio channel based on available resources;

- notifies the MEC server and the computing node of the established information exchange channel.

The second stage, where calculations and their verification are performed, includes the following steps (Fig. 3):

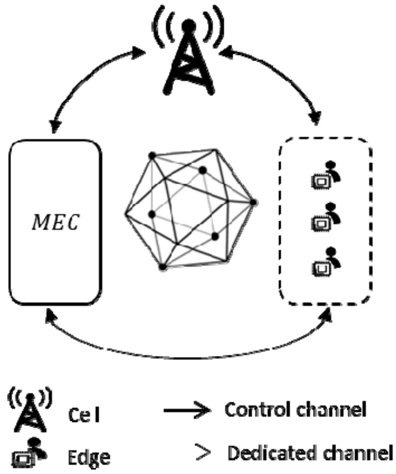

Fig. 3. The sequence of actions at the second stage of distributed computing — performing calculations and verifying results

1. The MEC server and the computing node establish a radio channel. The radio channel is formed based on the channel configuration parameters that each participant receives from the base station.

2. The MEC server and the computing node then perform a synchronization procedure.

3. Based on ETSI, the computing node makes an API call (Fig. 4) and sends a report to the base station after the computation is complete.

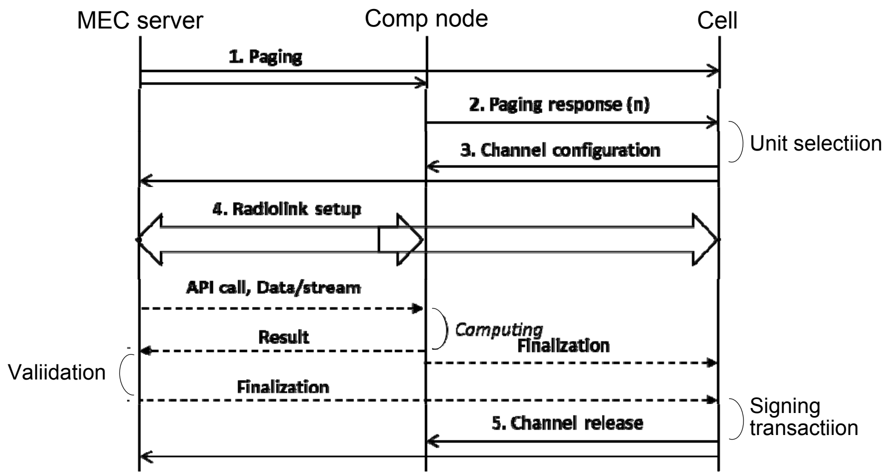

Fig. 4. The sequence of actions during an API call at the second stage

4. After verifying the result, the MEC server reports to the base station.

5. The reward for performing computations is calculated based on the complexity of the operation, execution time, and the amount of disk space consumed. To provide the compulsoriness of the workflow a blockchain technology is used. The participants are rewarded proportionally their fitness/ activities.

As mentioned above, during the second stage, the MEC server must check the calculations for correctness and errors, and assign a certain level of trust to each computing node. These procedures provided in the proposed method of data verification and error detection for the MEC system.

DATA VALIDATION AND ERROR FOUNDING METHOD FOR MEC SYSTEM

Each task that will be processed on the MEC server contains parts that can be performed independently. These parts are added to the task by software developers in the form of an API call. The results of such external computations carry the risks of computational errors and various types of attacks. In this paper, we propose a combined system for verifying the results of work, which includes the analysis of confidence levels and redundancy.

The proposed method for verifying the results of calculations and finding errors includes the following steps:

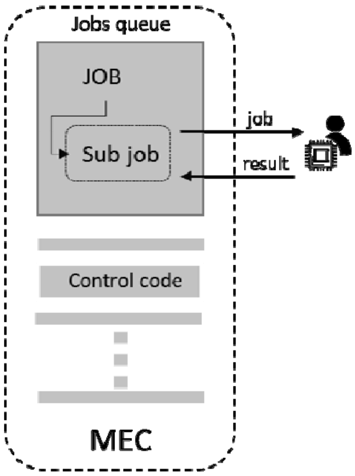

Fig. 5. Using a control code to check the correctness of calculations

1. The computing device of the MEC server generates an error control code in the form of a set of low-computing level functions.

The control code (Fig. 5) is an automatically created task with the same complexity, format and length of input data as the real task, with the only difference being that the MEC server knows the exact result, so it can be checked.

2. The MEC server distributes tasks between the MEC computing nodes with additional redundancy.

Redundancy (Fig. 6) additionally helps to avoid mistakes in computing even on trusted devices. MEC server will apply calculation results and grant rewards only after at least 51% current network nodes will present the same results.

3. The MEC server updates its trust level after the task is successfully completed (Fig. 7).

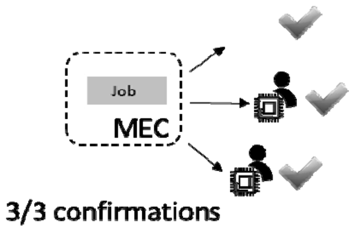

Fig 6. Use of additional redundancy to protect calculations from errors

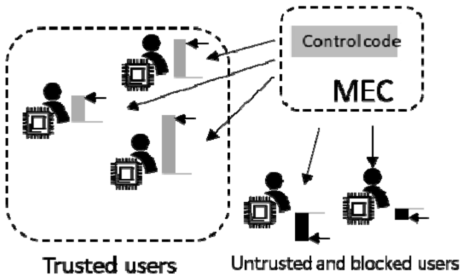

Fig. 7. Using the trust level to select computing nodes

Each MEC server has its own “trust level”, depended on the control code execution and the results of previously completed tasks. If the results of the check code execution are correct, the trust level (Fig. 7) for this computing node increases. Otherwise, if the device calculates the check code with errors, the level of trust decreases until the node is completely blocked.

ADVANTAGES OF THE PROPOSED APPROACH

The benefits of the proposed approach consist of two parts: the benefits to the user and the benefits to the mobile operator. For the user, the proposed approach provides:

- *Ease of setup:* common network identifiers (e.g. IMSI) or blockchain wallet number can be used as the user's device identifier.
- *Mobility*: MEC calculation can be done without bidding for mobile phones/locations.
- *High level of security*: The transaction is signed using blockchain technology. Mutual authentication allows verifying MEC server.
- *High reliability*: Data verification, mirroring and check code are used to protect against fraud and detect errors.
- Network operator benefits:
- *Network resource economy* (compare to existing solutions [9–18])*:* Dynamic map with node list (or another database / route table) does not need.
- *Low cost* (compare to existing solutions [9–18])*:* no additional hardware is required; a software upgrade can resolve this problem.
- *Easy UE selection and MEC load balancing:* base station make a decision based on the set of computing requirements.
- *Increasing Spectral Efficiency:* peer-to-peer communication release high load on cellular network.

CONCLUSION

The proposed approach for distributed edge computing in 5G allows identifying and authenticating MEC participants, allocating additional resources for MEC from the mobile network, including preparing point-to-point communications. The method also assigns computing nodes and balances the load of edge computing by modifying the messaging protocol between the base station and mobile devices.

The originality of the proposed approach is provided by two methods that are further improvements to the methods of load balancing, selection of computing nodes for edge computing in a 5G network, data verification and error detection for the MEC system. These methods are optimized according to the criteria of minimum network resources usage and have a time constraint.

The proposed approach allows the MEC server to verify the results of calculations and distribute data for computation according to the capacity of the computing nodes.

Implementation of this approach allows the service provider to save network resources and low cost of deployment. It also provides easy load balancing between computing nodes. This approach is more convenient for the user, as it does not require the creation of additional identifiers and provides a high level of security through the introduction of mutual authentication.

REFERENCES

1. F. Vhora, J. Gandhi, "A Comprehensive Survey on Mobile Edge Computing: Challenges, Tools, Applications," *2020 Fourth International Conference on Computing*

Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 49–55. doi: 10.1109/ICCMC48092.2020.ICCMC-0009

2. S.S.D. Ali, H. Ping Zhao, and H. Kim, "Mobile Edge Computing: A Promising Paradigm for Future Communication Systems," *TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018*, pp. 1183–1187. doi: 10.1109/TENCON.2018.8650169
3. L.A. Haibeh, M.C.E. Yagoub, and A. Jarray, "A Survey on Mobile Edge Computing Infrastructure: Design, Resource Management, and Optimization Approaches," in *IEEE Access*, vol. 10, pp. 27591–27610, 2022. doi: 10.1109/ACCESS.2022.3152787
4. D. Liu, A. Hafid, and L. Khoukhi, "Workload Balancing in Mobile Edge Computing for Internet of Things: A Population Game Approach," in *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1726–1739, 1 May-June 2022. doi: 10.1109/TNSE.2022.3150755
5. W. Zhang, G. Zhang, and S. Mao, "Joint Parallel Offloading and Load Balancing for Cooperative-MEC Systems With Delay Constraints," in *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4249–4263, April 2022. doi: 10.1109/TVT.2022.3143425
6. J. Ren et al., "An Efficient Two-Layer Task Offloading Scheme for MEC System with Multiple Services Providers," *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications, London, United Kingdom, 2022*, pp. 1519–1528. doi: 10.1109/INFOCOM48880.2022.9796843
7. W. Zhang, W. Fan, G. Zhang, and S. Mao, "Learning-based joint service caching and load balancing for MEC blockchain networks," in *China Communications*, vol. 20, no. 1, pp. 125–139, Jan. 2023. doi: 10.23919/JCC.2023.01.011
8. H. Wu, L. Chen, C. Shen, W. Wen, and J. Xu, "Online Geographical Load Balancing for Energy-Harvesting Mobile Edge Computing," *2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 2018*, pp. 1–6. doi: 10.1109/ICC.2018.8422299
9. "The Radio Access Network and its resource allocation method calculated based on mixing fog," Patent App. CN108243245A, 2020.
10. Debashish Purkayastha, Xavier De Foy, Robert G. Gazda, "Systems and methods to create slices at a cell edge to provide computing services," Patent App. WO2018089417A1.
11. "Resource allocation methods based on mobile edge calculations," Patent App. CN109041130A, 2021.
12. "A kind of data distribution method based on MEC auxiliary in 5G networks," Patent App. CN108174421A, 2020.
13. "A kind of edge calculations method for 5G super-intensive networking scenes," Patent App. CN107333267A, 2019.
14. "A kind of 5G method of mobile communication and system based on MEC and layering SDN," Patent App. CN107404733A, 2020.
15. John Juha Antero Rasanen, Pekka Kuure, "Method and apparatus for implementing mobile edge application session connectivity and mobility," US Patent App. US16/072,901 (US 2019/0045409 A1).
16. Mehran Moshfeghi, "System and method for discount deal referral and reward sharing," US Patent App. US13/484,261 (US20120316939A1).
17. Willem Jacobus van Niekerk, Marc van Niekerk, Brendon van Niekerk, Ernst Kleinhans, "Method and system for online redistribution of data and rewards," US Patent App. US15/464,934 (US20170255981A1).
18. Dennis Detwiller, Robert Sparks, "Rewarding Users for Sharing Digital Content," US Patent App. US13/106,716 (US20120290308A1).
19. Pavel Mach, Zdenek Becvar, *Mobile Edge Computing: A Survey on Architecture and Computation Offloading*. Available: https://arxiv.org/abs/1702.05309

20. Q.V. Pham et al., "A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art," *IEEE Access*, 2020. doi: https://doi.org/10.1109/ACCESS.2020.3001277

Received 02.02.2024

INFORMATION ON THE ARTICLE

Andrii A. Astrakhantsev, ORCID: 0000-0002-6664-3653, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: andrii.astrakhantsev@nure.ua

Larysa S. Globa, ORCID: 0000-0003-3231-3012, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: lgloba@its.kpi.ua

Oleksandr V. Fedorov, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

Dmytro V. Degtiarov, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

Yevgen O. Romanko, Interregional Academy of Personnel Management, Ukraine

Kyrylo A. Romanii, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

УДОСКОНАЛЕНИЙ ПІДХІД ДО ОРГАНІЗАЦІЇ МОБІЛЬНИХ ПЕРИФЕРІЙНИХ ОБЧИСЛЕНЬ У МЕРЕЖІ 5G / А.А. Астраханцев, Л.С. Глоба, О.В. Федоров, Д.В. Дегтярьов, Є.О. Романко, К.А. Романій

Анотація. Мобільні периферійні обчислення є важливим елементом забезпечення ефективності мережі 5G в цілому, оскільки дозволяють зберігати дані та виконувати обчислення на периферії мережі. В існуючих технічних рішеннях для систем зв'язку не в повному обсязі вирішені питання розподілу навантаження між обчислювальними вузлами; у більшості таких рішень не пропонується метод балансування обчислювального навантаження з контролем помилок у децентралізованій обчислювальній інфраструктурі з динамічно змінюваним набором обчислювальних вузлів. Метою дослідження є розроблення підходу до організації мобільних периферійних обчислень у мобільній мережі 5G, який би виконував перевірку справжності серверів розподілу та обчислювальних вузлів, керував процесом розподілу обчислювальних блоків, мав процедуру перевірки коректності розрахунків та враховував параметри обчислювальних вузлів під час розподілу. Для досягнення мети пропонується застосувати метод балансування навантаження та вибору обчислювальних вузлів для периферійних обчислень в мережі 5G, який дозволяє визначити наявні вузли та здійснити розподіл обчислювальних блоків між ними, а також забезпечити взаємну автентифікацію елементів інфраструктури з перевіркою даних та пошуком помилок для системи MEC, який дає змогу контролювати появу помилок під час обчислень, захищати сервер від некоректних даних. Указаний метод оптимізовано за критеріями мінімуму використовуваних ресурсів мережі і мінімального часу виконання обчислень. Такі вдосконалення дозволяють підвищити ефективність мобільних периферійних обчислень у 5G мережі.

Ключові слова: мережі 5G, мобільні периферійні обчислення, розподіл завдань, протокол обміну, балансування навантаженням, верифікація обчислень.