

INFORMATION TECHNOLOGY, COMPUTER SCIENCE AND MANAGEMENT ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ



UDC 519.873+ 519.876.2

<https://doi.org/10.23947/2687-1653-2023-23-1-85-94>

Original article



Two-Criteria Technique for the Resource-Saving Computing in the Fog and Edge Network Tiers

 Anna B Klimenko 

Institute of IT and Security Technologies, RSUH, 25, Kirovogradskaya St., Moscow, Russian Federation

✉ Anna.klimenko@mail.ru

Abstract

Introduction. At present, the concepts of fog and edge computing are used in a wide range of applications of various kinds. One of the key problems in the organization of computing in groups of mobile devices that make up the edge/fog layer is the mission assurance based on battery power availability. In this context, a lot of developments aimed at energy saving of device systems have been presented to date. However, one important aspect remains beyond the consideration of the problem of resource saving, namely, the issue of saving the residual resource of a computing device. The aim of this research is to formalize the workload distribution problem as two-criteria optimization problem, and to develop the basic solution technique.

Materials and Methods. Within the framework of this article, an approach to resource saving is proposed. It is based on the evaluation of two device criteria: battery life and residual resource of a computing device. The residual resource of a computing device can be estimated using the probability of failure-free operation of the device, or as the reciprocal of the failure rate, taking into account that the exponential law of failure distribution is used in the simulation. From this, a model of the problem of two-criteria optimization is formulated, taking into account the dynamics of the network topology in the process of performing a user mission. The topology dynamics is reflected in the model as a sequence of topologies, each of which corresponds to a certain period of time of the system operation.

Results. Based on the proposed model of the two-criteria optimization problem, a method was proposed for resource saving in the edge and foggy layers of the network. It reflected the specifics of the dynamic layers of the network, and also took into account the importance of the criteria for estimating the consumption of device resources. An experiment was conducted to evaluate the impact of the method of distributing tasks over a network cluster on the probability of failure-free operation of devices and on the average residual resource.

Discussion and Conclusions. The conducted experiment has demonstrated the feasibility of using the developed method, since the distribution of tasks among executing devices had a significant impact (up to 25 % according to the results of the experiment) on the average residual resource of a computing device.

Keywords: resource saving, calculation planning, fog computing, edge computing, optimization.

Acknowledgements. The author would like to thank the management of the Institute of IT and Security Technologies, Russian State University for the Humanities, for the assistance provided during the preparation of the project.

For citation. Klimenko AB. Two-Criteria Technique for the Resource-Saving Computing in the Fog and Edge Network Tiers. *Advanced Engineering Research (Rostov-on-Don)*. 2023;23(1):85–94. <https://doi.org/10.23947/2687-1653-2023-23-1-85-94>

Двухкритериальный метод обеспечения ресурсосбережения в краевом и туманном слоях сети

А.Б. Клименко 

Институт информационных наук и технологий безопасности РГГУ, Российская Федерация, г. Москва, ул. Кировоградская, д. 25

✉ Anna.klimenko@mail.ru

Аннотация

Введение. В настоящее время концепции туманных и краевых вычислений используются широким кругом приложений самой различной направленности. Одной из ключевых проблем организации вычислений в группах мобильных устройств, составляющих краевой/туманный слой, является обеспечение выполнения миссии на основе наличия заряда батареи. В связи с этим к настоящему времени представлено немало разработок, направленных на энергосбережение систем устройств. Однако очень важный аспект остается за рамками рассмотрения проблемы ресурсосбережения, а именно — вопрос сбережения остаточного ресурса вычислительного устройства. Целью данного исследования является формализация задачи распределения нагрузки как двухкритериальной задачи оптимизации и выбор базового метода ее решения.

Материалы и методы. В рамках данной статьи предлагается подход к ресурсосбережению на основе оценивания двух критериев устройств: ресурса батареи и остаточного ресурса вычислительного устройства. Остаточный ресурс вычислительного устройства может быть оценен при помощи значений вероятности безотказной работы устройства или как величина, обратная интенсивности отказов с учетом того, что при моделировании используется экспоненциальный закон распределения отказов. На основе этого сформулирована модель задачи двухкритериальной оптимизации с учетом динамики топологии сети в процессе выполнения пользовательской миссии. Динамика топологии отражена в модели как последовательность топологий, каждая из которых соответствует определенному отрезку времени функционирования системы.

Результаты исследования. На основании представленной модели задачи двухкритериальной оптимизации предложен метод обеспечения ресурсосбережения в краевом и туманном слое сети, отражающий специфику динамических слоев сети, а также учитывающий важность критериев оценивания расхода ресурсов устройств. Проведен эксперимент, позволяющий оценить влияние способа распределения задач по сетевому кластеру на вероятность безотказной работы устройств и на средний остаточный ресурс.

Обсуждение и заключения. Проведенный эксперимент демонстрирует целесообразность применения разработанного метода, поскольку распределение задач по исполняющим устройствам оказывает существенное влияние (до 25 % по итогам эксперимента) на средний остаточный ресурс вычислительного устройства.

Ключевые слова: ресурсосбережение, планирование вычислений, туманные вычисления, краевые вычисления, оптимизация.

Благодарности. Автор выражает признательность руководству Института информационных наук и технологий безопасности Российского государственного гуманитарного университета за помощь, оказанную в процессе подготовки проекта.

Для цитирования. Клименко А.Б. Двухкритериальный метод обеспечения ресурсосбережения в краевом и туманном слоях сети. *Advanced Engineering Research (Rostov-on-Don)*. 2023;23(1):85–94. <https://doi.org/10.23947/2687-1653-2023-23-1-85-94>

Introduction. Currently, applications using edge and fog network segments are widely used [1–5]. Distributed computing performed on the nodes of these segments has significant difference from distributed computing performed

in cloud structures, namely, relatively high topology dynamics and, in addition, the presence of certain distances between the nodes used. This results in the need to take this parameter into account when modeling the time of information exchanges between executable tasks. The existence of topology dynamics is explained by the fact that network edge devices are, as a rule, user devices (including all kinds of sensors, smartphones, laptops, etc.) that belong to certain people and can be moved, turned off, access to data/resources may be prohibited. At the same time, all of them depend on the presence of a battery charge (power source) [6, 7]. The fog layer of the network is less dynamic. However, it is also characterized by all of the above to one degree or another: the occurrence of a certain distance between nodes with the presence of transit sections of the network, the possible mobility of the nodes themselves, e.g., in case of data processing by groups of mobile devices. An interesting example in this sense are groups of low-orbit satellites that are currently used for data processing, but the nodes are displaced relative to the Earth's surface at a fairly high speed (access time is on the order of several minutes) and, therefore, the question arises about the routing and transmission of processing results to ground stations [8].

As part of the issues of resource saving of devices used in the dynamic layers of the network, a number of works have been published [9–11], in which attention is focused on energy saving. As a rule, the model assumes the dependence of the consumed energy of the device on its workload [12]. At the same time, such an important factor as saving the residual computing resource of the device remained outside the focus of research [13].

The residual computing resource of a device is a value that is closely related to such reliability characteristics as the probability of failure-free operation and gamma-percentage time to failure. Also, the average residual life, taking into account the use of the exponential law of failure rate distribution, is characterized by the following expression:

$$R(\tau) = \frac{1}{P(\tau)} \int_{\tau}^{\infty} P(x) dx, \quad (1)$$

where $P(x)$ — probability of faultless operation (PFO) of the object during time x .

For the exponential resource distribution law, the PFO is defined as:

$$P(x) = e^{-\lambda x}, \quad (2)$$

where λ — hazard rate, $\lambda > 0$.

Then, according to formula (1), the average residual life is determined from the following formula:

$$R(\tau) = R = \frac{1}{\lambda}. \quad (3)$$

At the same time, there are works devoted to the relationship between the values of PFO, gamma-percentage time to failure, and the average residual resource on the temperature of the computing element of the device, which, in turn, is described as a function of the device load [14–16]. The problem of device resource saving can also be posed as the problem of maximizing the average residual resource of the device.

Thus, the problem of resource saving is considered in recent works either as a task of minimizing the power consumption of devices, or as a task of saving the residual computing resource.

However, under present-day conditions, it is reasonable to talk about two-criteria optimization of resource consumption of devices for the following reasons:

- electricity is a renewable but critical resource for the mission of the device, whether it is a sensor of an information and control system or a controlled mobile device;
- residual computing resource determines the duration of the device's expedient operation, and the operation time, respectively, is the sum of the missions performed by the device.

Therefore, the objective of this study was to develop a resource saving method for devices of dynamic network layers, within which the assessment would be carried out on the basis of two criteria — power consumption and computing resource consumption of the device.

Materials and Methods. Let us define the concept of “mission of a group of devices” as a predetermined sequence of solving by a group of devices a complex of computational tasks related to information exchanges. The mission of a group of devices can be described by an acyclic graph, whose vertices will be weighted by the complexity of the tasks, and the arcs, respectively, will determine the limits of the sequence and the amount of data transmitted by the tasks. Let the mission be described by graph $W = \{w_l, z_l, I\}$, where w_l — complexity of the task, z_l — proportion of the task completed by the time of the topology change, I — matrix of data volumes transferred between tasks. A group of devices in a dynamic network layer is described by a set of graphs, each of which defines the topology of the network at time t_i . We assume that the mission execution time includes a certain time interval with discrete instants of time $[t_0, t_1, \dots, t_k]$, each of which corresponds to the topology graph $G_i = \{p_{ij}, e_{ij}, R_{ij}, H\}$, $j=1 \dots m$, p_{ij} — performance of the j -th node of the i -th topology, e_{ij} — energy resource of the j -th node of the i -th topology, R_{ij} — average residual computing resource of the j -th node of the i -th topology, m — number of devices in a group, H — matrix of network connections between nodes. We also assume that at each subsequent time t_{i+1} , for graphs G_{i+1} and G_i , the following is true: at least one of the vertices belonging to G_i , will be contained in G_{i+1} , which is required to continue the mission.

We also assume that if the topology graph changes at time t_{i+1} , it is possible to reassign tasks that were performed in the i -th topology on the nodes that ceased to exist at $i+1$ time, to any of the available topology nodes $i+1$. At the same time, we make the assumption that the tasks transferred to new nodes start to be executed at the same place where their execution was interrupted.

Estimation of the residual computational resource will be implemented as follows: to estimate the PFO of the device of the fog layer, we will use the expressions proposed in [17, 18]:

$$\lambda = \lambda_0 \cdot 2^{\frac{kD}{10}}, \quad (4)$$

where D — calculator load (in fractions).

$$D = \frac{w_i}{p_j \cdot t_{\text{constraint}}}, \quad (5)$$

where w_i — complexity of the work performed.

Accordingly, PFO (t) can be estimated, as well as the average residual resource:

$$P(t) = e^{-\lambda t} = e^{-\lambda_0 \cdot t \cdot 2^{\frac{kD}{10}}} = e^{-\lambda_0 \cdot t \cdot 2^{\frac{k \cdot w_i}{10 \cdot p_j \cdot t_{\text{constraint}}}}}, \quad (6)$$

where $t_{\text{constraint}}$ — time limit for which the task must be completed by the node to meet the mission time limit t_m .

When estimating the residual energy resource, we also assume that the energy consumption is proportional to the workload of the computing node, and there are no more energy costs:

$$E_{jm} = E_{j0} - \xi D, \quad (7)$$

where E_{j0} — initial level of energy available, E_{jm} — residual level, ξ — coefficient expressing the relationship between the load of the computing element and its energy consumption.

The result of the operation of the two-criteria method for ensuring resource saving will be a sequence of distributions of still unsolved mission tasks over time $[t_0, t_{i-1}]$ by topology G_i taking into account the fact that some of the tasks that are in the process of being solved on the nodes inherited from topology G_{i-1} , remain at their places and are not transferred: $A = \{A_i\}$;

$$A_i = \begin{vmatrix} t_{11} & \dots & \dots \\ \dots & t_{ij} & \dots \\ \dots & \dots & t_{ik} \end{vmatrix}. \quad (8)$$

Matrix A describes the distribution of unsolved problems for which $z_l < 1$. At the same time, the number of nodes corresponds to new topology G_i . Nodes inherited from G_{i-1} , are renumbered and occupy indices starting from 1.

Since the mission may involve more than one topology, we will decompose the problem as follows: we will use the “greedy” policy and assume that a satisfying solution (minimization of resource consumption, estimated by two chosen criteria) can be obtained, if for each topology, we choose the best solution.

That is, the task model will have the following form: $\forall g_i, z_i < 1, t_i > 0$ to get such A_i , that $R(t_m)$ $R_j(t_m) \rightarrow \max, e_j(t_m) \rightarrow \max$, при $\max(A_{ijk}) \leq t_m$.

Consider the possible priority of the formulated objective functions. For each individual topology, it is required to minimize both the power consumption and the consumption of computing resources. In some special cases, the coincidence of optimization goals for two criteria can be achieved. Specifically, the task is distributed to the node in such a way that it will provide it with the minimum workload of the computational element and at the same time the minimum power consumption.

However, this situation is not always possible. For example, there is a node loaded less than other calculations, but at the same time its energy resource is almost exhausted. By assigning a task to it, we jeopardize the mission (or it is necessary to change topologies as the supply of electricity runs out), and by assigning a task to a more loaded node, we worsen the value of the residual computing resource criterion. The opposite situation can also occur: a node with a large residual power resource may have a low residual computing resource as a result of participating in previous device missions.

Therefore, it is advisable to rank the objective function (OF) in terms of importance, on the basis of which to select a further solution method.

For critical missions, the battery life of the device plays a leading role [19–21]. Accordingly, the OF that determines the residual computing resource can be reduced to an additional constraint. Uncompleted mission tasks will be redistributed depending on the remaining battery charge.

It should be noted that the above is also true for missions that do not have critical execution. In the event that the device loses battery power without completing the mission, it may be lost. Then the optimization of the computing resource will become meaningless.

On the other hand, by reducing the residual computational resource to a limitation, the solution may not be obtained at all due to the lack of available nodes whose characteristics would satisfy it.

Research Results. Multicriteria optimization problems, as a rule, are reduced to solving one or more single-criteria problems. In this case, it is more expedient to leave one OF, and present the second one as a constraint.

For the system under consideration, the fulfillment of the mission is critical for each node; therefore, we transform the OF for energy consumption into a restriction, the implementation of which will cut off unacceptable options for distributing tasks among the nodes. Further, within the framework of the resulting set of nodes, the distribution will occur with minimization of the consumption of the residual computing resource.

Let us describe the main stages of the two-criteria method for ensuring resource saving as the basic result of the study:

- on the existing G_i -topology, select nodes that have sufficient energy resources to complete the mission without changing the topology;

– distribute mission tasks among selected nodes in such a way as to minimize the consumption of computing resources for idle nodes, to which unsolved tasks will be attached.

It is possible to rank the nodes obtained at the first step of the method in descending order of the availability of an energy resource. But then, by distributing tasks primarily among nodes with the maximum amount of energy resource, there may be a deterioration in the values of the residual computing resource criterion.

Let us illustrate the dependence of the PFO values of devices on the selection of a node for solving the problem (in this experiment, we compare the states of the nodes under the transmission of incoming data and during their processing).

The topology is shown in Figure 1.

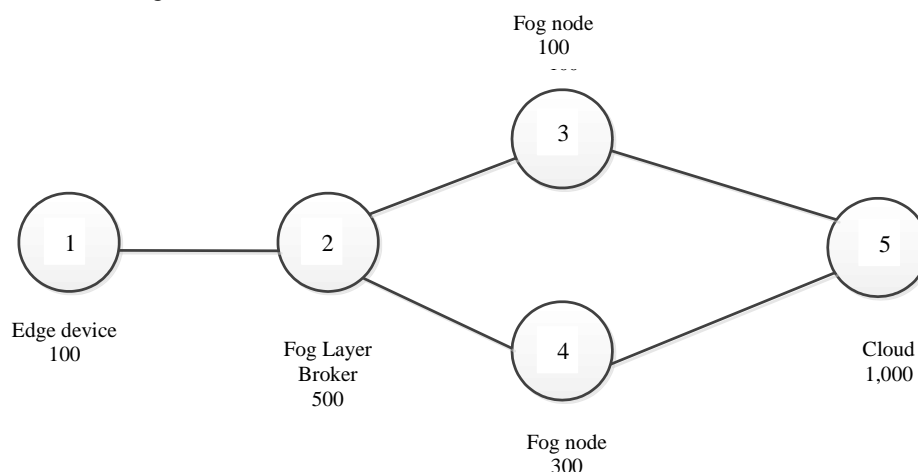


Fig. 1. Experimental fragment of network topology

The experiment was carried out for 10-time intervals, each of which was 100 hours with the following parameter values: Wreceive (amount of data received by the task) — 500 units, Wsend (amount of data sent by the task) — 100 units, Wprocess (data processing complexity) — 150 units, Tdecl (time allotted for solving the task) — 50 units.

The following are graphs of PFO nodes depending on whether the node processes data or only transmits it (Fig. 2–5).

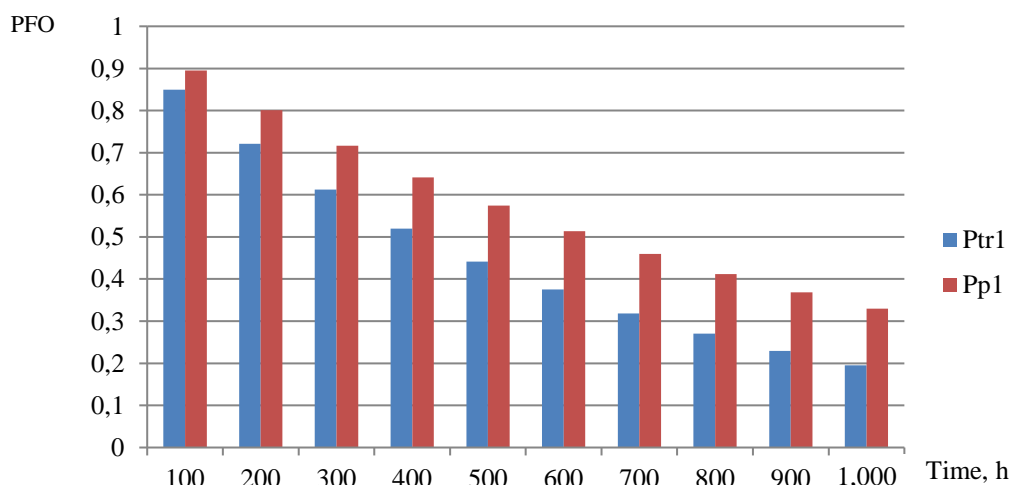


Fig. 2. Comparison of PFO values for node 1 under data processing (Pp1) and when transferring data (Ptr1)

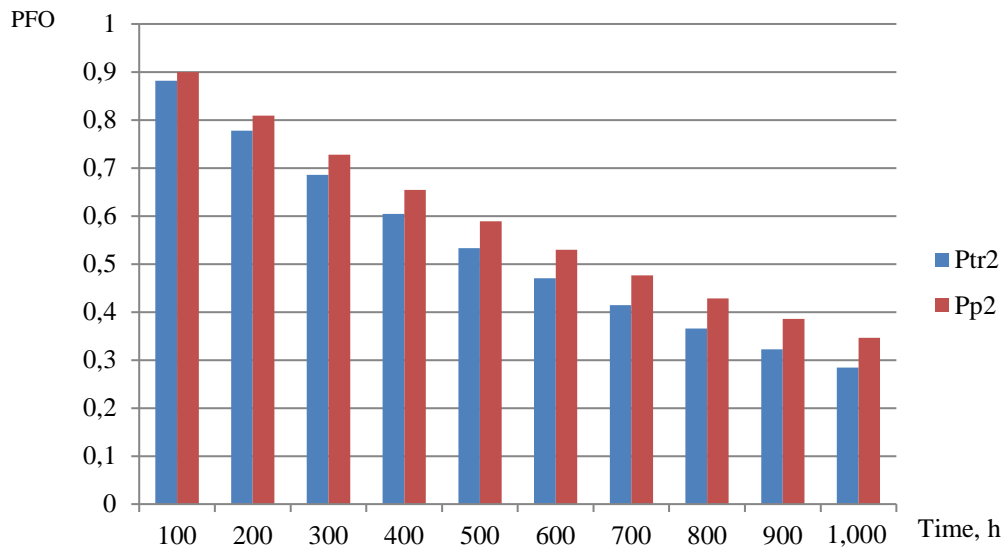


Fig. 3. Comparison of PFO values for node 2 under data processing (Pp2) and when transferring data (Ptr2)

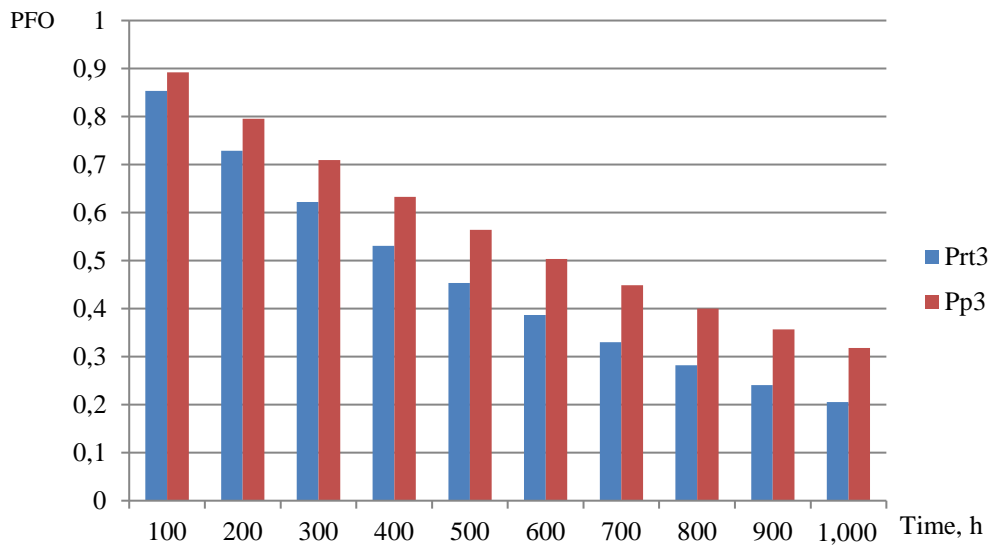


Fig. 4. Comparison of PFO values for node 3 under data processing (Pp3) and when transferring data (Ptr3)

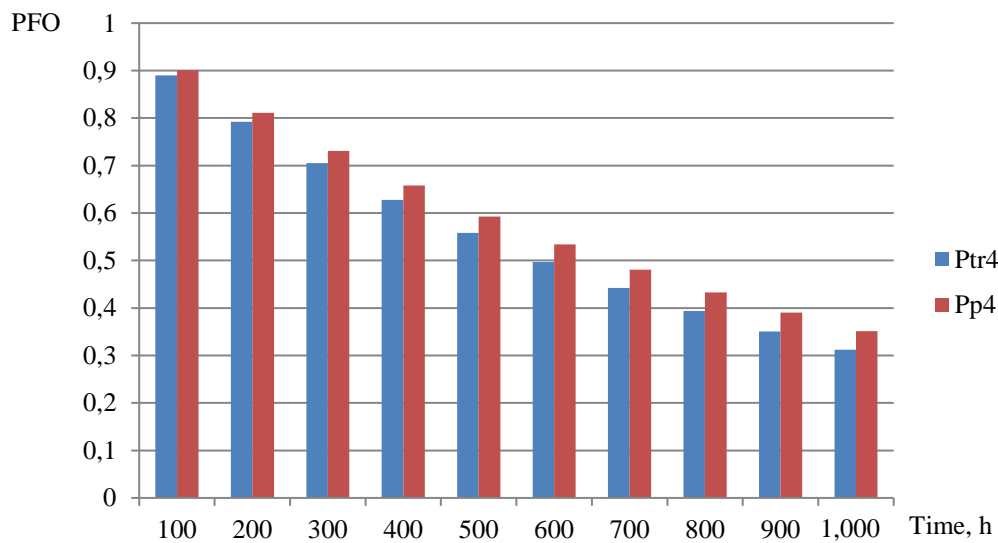


Fig. 5. Comparison of PFO values for node 4 under data processing (Pp4) and when transferring data (Ptr4)

Further, Figure 6 shows comparison of differences in the values of the average residual computing resource.

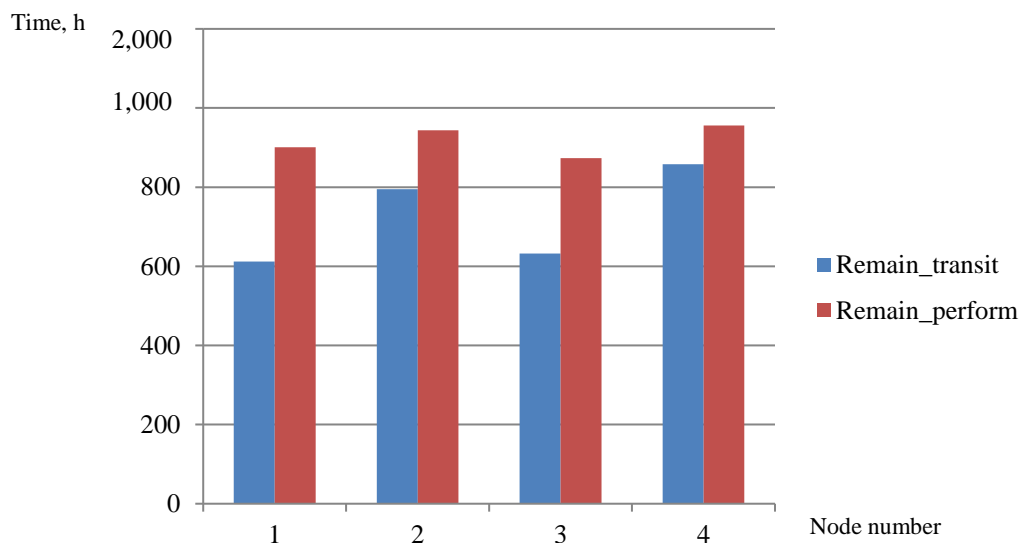


Fig. 6. Comparison of average residual resource values for nodes 1–4 during data transit (Remain_transit) and under data processing (Remain_perform)

Discussion and Conclusions. The article proposes a two-criteria method for ensuring resource saving in the edge and fog layers of the network. It differs from those previously presented in the subject area of organizing distributed computing in the dynamic layers of the network in the following. The proposed method takes into account both the state of the battery of the device (which is critical for the mission) and the residual resource of the computing device (what affects the duration of the appropriate use of the device outside the mission).

The developed method is based on the formulated problem of two-criteria optimization, which is then proposed to be reduced to a single-criteria one through converting the objective function of the energy resource balance into a constraint.

The nodes that are not suitable for the battery level are cut off in advance. Then there is a distribution of tasks according to the criterion of the residual computing resource.

The conducted experiment makes it possible to compare the results of task distribution over nodes in the presence of information interactions between tasks. It is clearly shown that the method of task distribution affects significantly the value of the residual resource of a computing device (up to 25 %). Thus, it is possible to improve the resource consumption indicators of devices involved in calculations in the edge and fog layers of the network.

References

1. Mehbub Alam, Nurzaman Ahmed, Rakesh Matam, et al. *L3Fog: Fog Node Selection and Task Offloading Framework for Mobile IoT*. In: Proc. IEEE Conference on Computer Communications Workshops — IEEE INFOCOM 2022. P. 1–6. [10.1109/INFOCOMWKSHPS54753.2022.9798118](https://doi.org/10.1109/INFOCOMWKSHPS54753.2022.9798118)
2. Abdelfettah Maatoug, Ghalem Belalem, Saïd Mahmoudi. A Location-Based Fog Computing Optimization of Energy Management in Smart Buildings: DEVS Modeling and Design of Connected Objects. *Frontiers of Computer Science*. 2023;17:172501. <https://doi.org/10.1007/s11704-021-0375-z>
3. Abohamama AS, Amir el Ghamry, Eslam Hamouda. Real-Time Task Scheduling Algorithm for IoT-Based Applications in the Cloud–Fog Environment. *Journal of Network and Systems Management*. 2022;30(4):54. <https://doi.org/10.1007/s10922-022-09664-6>
4. Saad Ahmad Khan, Muhammad Abdullah, Waheed Iqbal, et al. Efficient Job Placement Using Two-Way Offloading Technique over Fog-Cloud Architectures. *Cluster Computing*. 2022;154:1–19. [10.1007/s10586-022-03750-9](https://doi.org/10.1007/s10586-022-03750-9)

5. Rajasi Gore, Shashwati Banerjea, Neeraj Tyagi. A Heterogeneous Soft-Hard Fusion Framework on Fog Based Private SaS Model for Smart Monitoring of Public Restrooms. *Journal of Ambient Intelligence and Humanized Computing*. 2022. Vol. 63. <https://doi.org/10.1007/s12652-022-04401-y>
6. Haibo Wang, Hongli Xu, He Huang, et al. Robust Task Offloading in Dynamic Edge Computing. *IEEE Transactions on Mobile Computing*. 2021;22:500–514. <https://doi.org/10.1109/TMC.2021.3068748>
7. Enlu Liu, Xiaoheng Deng, Zhi Cao, et al. *Design and Evaluation of a Prediction-Based Dynamic Edge Computing System*. In: Proc. 2018 IEEE Global Communications Conference (GLOBECOM). <https://doi.org/10.1109/GLOCOM.2018.8647199>
8. Fraire JA, Gerstacker C, Hermanns H, et al. On the Scalability of Battery-Aware Contact Plan Design for LEO Satellite Constellations. *International Journal of Satellite Communications and Networking*. 2020;39:193–204. <https://doi.org/10.1002/sat.1374>
9. Yung-Ting Chuang, Chiu-Shun Hsiang. A Popularity-Aware and Energy-Efficient Offloading Mechanism in Fog Computing. *The Journal of Supercomputing*. 2022;78:1–24. <https://doi.org/10.1007/s11227-022-04626-w>
10. Ida Syafiza Binti, Taizir EH El-Gorashi, Mohamed OI Musa, et al. Energy Efficient Fog Based Healthcare Monitoring Infrastructure. *IEEE Access*. 2020;8:197828–197852. <https://doi.org/10.1109/ACCESS.2020.3033555>
11. Fatemah S Behbehani, Taisir El-Gorashi, Jaafar MH Elmirghani. Power Minimization in Vehicular Cloud Architecture. [arXiv:2102.09011](https://arxiv.org/abs/2102.09011) 2021. <https://doi.org/10.48550/arXiv.2102.09011>
12. Shilpa B Kodli, Sujatha Terdal. Survey on Energy Efficient-Load Balancing in Cloud. *International Journal of Computer Applications*. 2022;184(25):15–24. <http://dx.doi.org/10.5120/ijca2022922301>
13. Fudukhin AV. Predicting the Reliability of Electronic Devices after Long-Term Storage. *Mathematical Machines and Systems*. 2004;4:164–170. (In Russ.)
14. Klimenko AB, Melnik EV. A Method of Improving the Reliability of the Nodes Containing Ledger Replicas. In book: Silhavy R, Silhavy P, Prokopova Z. (eds.) *Software Engineering Application in Informatics*. 2021. P. 584–592. https://doi.org/10.1007/978-3-030-90318-3_47
15. Klimenko A, Melnik E. Information and Control Systems with Distributed Ledger Usage: A Reliability Issue. In book: Radek Silhavy (ed.) *Artificial Intelligence in Intelligent Systems*. Springer; 2021. P. 133–144. https://doi.org/10.1007/978-3-030-77445-5_12
16. Melnik E, Safronenkova I, Kapustyan S. *The Efficiency Improvement of Robots Group Operation by Means of Workload Relocation*. In: Proc. Int. Conf. on Interactive Collaborative Robotics. 2021;450:126–137. https://doi.org/10.1007/978-3-030-87725-5_11
17. Melnik EV, Gorelova GV. About Effect of Processor Computational Load Balancing in Highly Reliable Distributed Information and Control Systems, Research Planning. *Mechatronics, Automation, Control*. 2012;11:29–35.
18. Kalyaev IA, Melnik EV. *Decentralized Computer Control Systems*. Rostov-on-Don: YuNTS RAN Publ.; 2011. 196 p. (In Russ.)
19. Preedipat Sattayasoonthorn, Jackrit Suthakorn. *Battery Management for Rescue Robot Operation*. In: Proc. 2016 IEEE Int. Conf. on Robotics and Biomimetics (ROBIO). 2016. P. 1227–1232. <https://doi.org/10.1109/ROBIO.2016.7866493>
20. Vinay Jadhav, Surendra Bhosale. *Battery Management System for Drones*. In: Proc. Int. Conf. on Electrical Electronics and Data Communication. 2022. URL: https://www.researchgate.net/publication/362889754_Battery_Management_System_for_Drones (accessed: 10.12.2022).

21. Wei Liu, Tobias Placke, Chau KT. Overview of Batteries and Battery Management for Electric Vehicles. *Energy Reports*. 2022;8:4058–4084. <https://doi.org/10.1016/j.egy.2022.03.016>

About the Author:

Anna B Klimenko, associated professor of the Fundamental and Applied Mathematics Department, Institute of IT and Security Technologies, RSUH (25, Kirovogradskaya St., Moscow, 117534, RF), Cand.Sci. (Engineering), [ScopusID](#), [ORCID](#), anna_klimenko@mail.ru

Received 23.12.2022

Revised 20.01.2023

Accepted 21.01.2023

Conflict of interest statement

The author does not have any conflict of interest.

The author has read and approved the final manuscript.

Об авторе:

Анна Борисовна Клименко, доцент кафедры «Фундаментальная и прикладная математика» Института информационных наук и технологий безопасности Российского государственного гуманитарного университета (117534, Москва, ул. Кировоградская, д. 25 корп. 2), кандидат технических наук, [ScopusID](#), [ORCID](#), anna_klimenko@mail.ru

Поступила в редакцию 23.12.2022

Поступила после рецензирования 20.01.2023

Принята к публикации 23.01.2023

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Автор прочитал и одобрил окончательный вариант рукописи.