

UDC 004.942

Doi: 10.31772/2587-6066-2020-21-4-492-498

For citation: Ontuzheva G. A. Models and methods of optimal control of software and technical configuration of heterogeneous distributed information processing systems. *Siberian Journal of Science and Technology*. 2020, Vol. 21, No. 4, P. 492–498. Doi: 10.31772/2587-6066-2020-21-4-492-498

Для цитирования: Онтужева Г. А. Модели и методы оптимального управления программно-технической конфигурацией гетерогенных распределенных систем обработки информации // Сибирский журнал науки и технологий. 2020. Т. 21, № 4. С. 492–498. Doi: 10.31772/2587-6066-2020-21-4-492-498

MODELS AND METHODS OF OPTIMAL CONTROL OF SOFTWARE AND TECHNICAL CONFIGURATION OF HETEROGENEOUS DISTRIBUTED INFORMATION PROCESSING SYSTEMS

G. A. Ontuzheva

Siberian Federal University
79, Svobodny Av., Krasnoyarsk, 660041, Russian Federation
E-mail: gontuzheva@sfu-kras.ru

The article discusses formalization of the problem of heterogeneous distributed information processing systems (HDIPS) software and hardware configuration management. A formal description of possible optimality criteria for the HDIPS software and hardware configuration is given. The HDIPS model in terms of queuing theory is proposed. The problem of allocating the HDIPS computational resources is formulated as a transport problem according to time criterion with atomic needs. The algorithm for solving this problem is proposed and the boundaries of its applicability to the HDIPS are determined. To meet the selected optimality criterion, the analysis of the HDIPS software and hardware configuration applying its formal model, using the queuing theory methods is presented. HDIPS is presented as a queuing network, where each computing node and route control unit is a mass service system. The problem of computing resource allocation in HDIPS is presented as a transport problem according to the time criterion with atomic needs. The least time algorithm for indivisible needs takes into account the indivisibility condition.

Keywords: distributed information processing systems, transport problem, queuing systems, software and hardware configuration, management of software and hardware resources, management optimization.

МОДЕЛИ И МЕТОДЫ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ПРОГРАММНО-ТЕХНИЧЕСКОЙ КОНФИГУРАЦИЕЙ ГЕТЕРОГЕННЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМ ОБРАБОТКИ ИНФОРМАЦИИ

Г. А. Онтужева

Сибирский федеральный университет
Российская Федерация, 660041, г. Красноярск, просп. Свободный, 79
E-mail: gontuzheva@sfu-kras.ru

В статье рассматривается формализация задачи управления программно-технической конфигурацией гетерогенных распределенных систем обработки информации (ГРСОИ). Дано формальное описание возможных критериев оптимальности программно-технической конфигурации ГРСОИ. Предложена модель ГРСОИ в терминах теории массового обслуживания. Задача распределения вычислительных ресурсов ГРСОИ сформулирована в виде транспортной задачи по критерию времени с атомарными потребностями. Предложен алгоритм решения данной задачи и определены границы его применимости в ГРСОИ. Для достижения выбранного критерия оптимальности приведен анализ программно-технической конфигурации ГРСОИ при помощи ее формальной модели с применением методов теории массового обслуживания. ГРСОИ представлена как сеть массового обслуживания, где каждый вычислительный узел и блок управления маршрутом является системой массового обслуживания. Задача распределения вычислительного ресурса в ГРСОИ представляется как транспортная задача по критерию времени с атомарными потребностями. С учетом условия атомарности был разработан алгоритм наименьшего времени для атомарных заявок.

Ключевые слова: распределенная системы обработки информации, транспортная задача, системы массового обслуживания, программно-техническая конфигурация, управление программно-техническими ресурсами, оптимизация управления.

Introduction. Heterogeneous distributed information processing systems (HDIPS) are information processing systems that are characterized by territorial distribution, a variety of software and hardware components and a heterogeneous nature of tasks being processed [1; 2]. Such systems are used in areas where it is necessary to receive and process primary data of a various nature in an automatic or automated mode. They combine computational nodes (CN) and data sources of various types, which make it possible to carry out the entire computation process in the system, from obtaining raw operational data obtained from one or several sources to delivering final information to decision-makers.

One of the HDIPS features as a class of systems is a heterogeneous nature of tasks simultaneously solved in the system. They may require various software and hardware resources, which increases the complexity of the most efficient software and hardware configuration choice. Due to the complexity of the HDIPS, a decision-making process for software and hardware configuration management is associated with a large amount of uncertainty, making decision-making on the design and modernization of the HDIPS software and hardware configuration laborious, and increases error probability. The number of HDIPS software and hardware configuration elements and permissible ways of connecting them into various structures, which perform computational functions, are great.

The more components a HDIPS contains, the higher the complexity of the interaction between them is, therefore special tools are needed to work with such a large amount of information.

Existing approaches are either intended for universal computing systems and do not take into account heterogeneity [3–7], or do not imply the possibility of changing the software and hardware configuration [8–10].

The inter-agency nature of some HDIPSs also complicates the system management. For example, in the event that a HDIPS was formed due to the merger of several departments or divisions, it is difficult to see the system “from above” without special tools, to assess its potential and the way to optimize the combined computing resource use by shifting from independent problem solving “old” subsystems to a shared computing space.

The aim of this work is to formalize the problem of HDIPS software and hardware configuration management. To achieve this goal, formalization of possible optimality criteria of HDIPS software and hardware configuration was carried out, a HDIPS model in terms of queuing theory was proposed, the problem of allocating HDIPS computing resources was formulated as a transport problem by the time criterion with atomic needs, the algorithm for its solution was proposed and its applicability boundaries were determined.

Optimality criteria for HDIPS software and hardware configuration. Software and hardware configuration is a set of functional parts of an information processing system, their software and connections between them, due to the main technical characteristics of these functional parts, as well as the requirements of the tasks to be solved [11].

The problem of HDIPS optimal configuration choice is the choice of such a set of CN P , which provides an acceptable level of the optimality criterion J for solving a set of computational problems E . In this case, the optimality criterion may differ depending on the purpose of a particular HDIPS [12]. Such criteria can be as follows:

1. Minimizing the CN utilization factor average value:

$$J^1 = \min(\overline{\text{utilasid}}).$$

In this case, we can assume that the computational load is evenly distributed and there are no overloaded nodes

2. Minimizing the total time on problem solving in the system

$$J^2 = \min\left(\sum_i^n T(e^i)\right),$$

where T – time spent on problem solving e^i , n – number of problems, calculated at the time of change. HDIPS can be used in areas where decision-making time is critical, in which case minimizing a problem processing time is more important than uniform load distribution.

3. Minimizing the probability of returning problem with CN due to a lack of computing resources:

$$J^3 = \min\left(\sum_i^n P_{return}^i\right).$$

In case of an incorrect combination of computational load, a number of CNs and algorithms for distributing the computing resource in the HDIPS, situations are possible when the task arrives at the CN, which does not have enough free resources to process it; in this case the task is returned to the routing agent. The likelihood of such returns must be minimized, since they indicate the non-optimal configuration of the HDIPS and increase the time spent on tasks in the computing system.

The choice of the efficiency criterion for the projected computing system is the first step in solving the problem of optimal HDIPS software and hardware configuration choice. To achieve the selected optimality criterion, it is proposed to analyze the HDIPS software and hardware configuration using its formal model, applying the queuing theory methods.

Representation of HDIPS in the form of a queuing network. HDIPS can be represented as a queuing network, where each CN and route control unit is a mass service system (QS). By a queuing network we mean a set of interconnected servicing devices with queues (queuing systems), in which requests pass from one device to another with a certain probability [13]. Each CN is represented as a multichannel QS without a queue, which returns a request for calculating a task to a routing agent if there is not enough computational resource for its execution.

Queues are accumulators for routing agents, while the probability of task transition to a specific CN is determined by the routing algorithm operation. A general scheme of the queuing network HDIPS is shown in fig. 1

The QS scheme of the CN is shown in fig. 2. Unserviced requests can arise in the QS of the CN in the event that upon the request receipt for the CN there is not enough resource to process it.

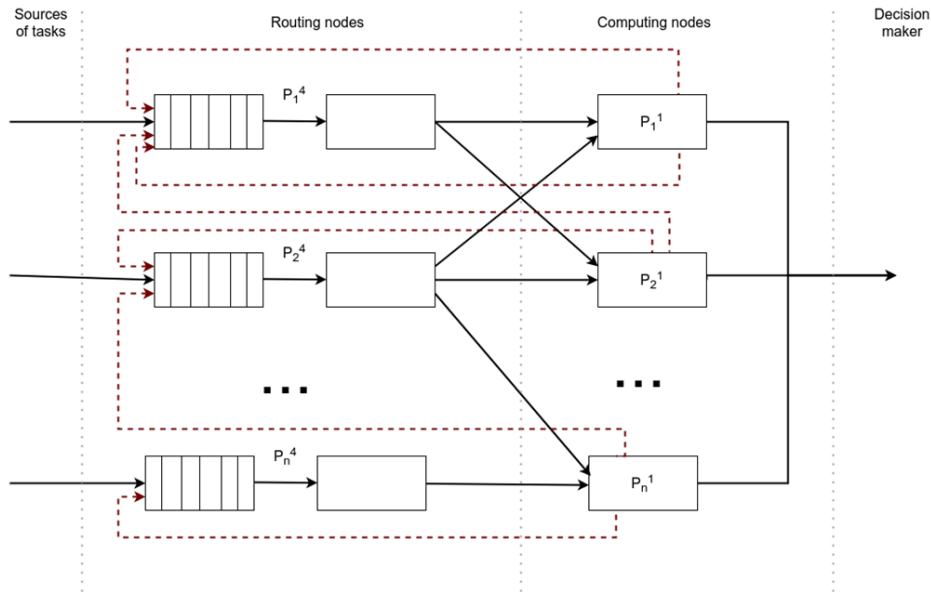


Fig. 1. A general scheme of the queuing network HDIPS

Рис. 1. Общая схема сети массового обслуживания ГРСОИ

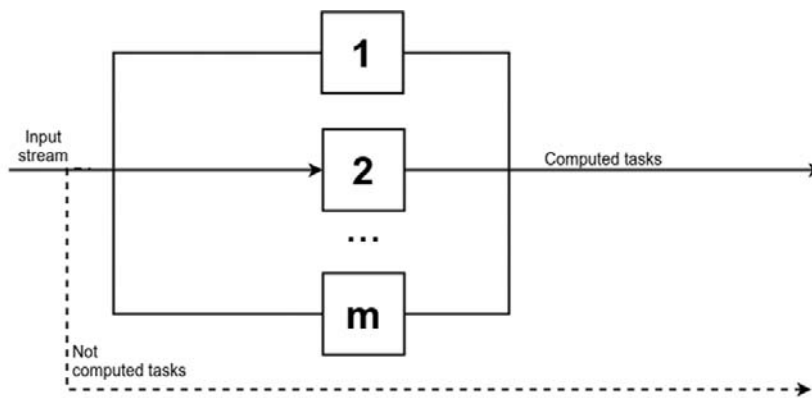


Fig. 2. The scheme of the queuing system computing node

Рис. 2. Схема системы массового обслуживания вычислительного узла

In this case, the probabilities of the task transition from the router p_i^4 to the CN p_j^1 is defined as the product request probability sending to this node in accordance with the routing algorithm operation and availability of the necessary software coefficient a_{task}^i in the software configuration of the control node a_{server}^j .

$$p_{ij} = p_{ij}^{routing} k_{ij},$$

where

$$k_{ij} = \begin{cases} 1, & \text{if CN } j \text{ is provided with required software,} \\ 0, & \text{if CN } j \text{ is not provided with required software.} \end{cases}$$

For each task subset requiring the same software set, only a CNs subset will be available that satisfies the condition of suitable software availability.

Thus, for a task from the A_{task}^k set, only a part of the queuing network elements will be available; therefore, it is advisable to calculate the average time spent on tasks in the system using the weighted average time spent in the

task system for each type of software from A_{task} , subsets, where the weight will be the probability of the task appearance, requiring A_{task}^k , software in the system. Thus, it is possible to represent the efficiency criterion J^2 using the queuing theory instruments:

$$J^2 = \bar{T} = \frac{1}{\Lambda} \sum_{j=1}^N w_i \lambda_j \bar{t}_{\Sigma_j},$$

where Λ – total intensity of network input streams, \bar{t}_{Σ_j} – average time spent by a task in j QS, λ_j – input flow rate j QS, w_i – weight of i task. To achieve the selected optimality criterion, in addition to enumerating different permissible combinations of software and hardware configurations of the system individual elements, it is necessary to determine the optimal algorithm for requests distribution for computations to computational nodes.

Transport task by the time criterion with atomic needs. The function that reflects the total time of data processing

problems at any finite time interval is presented in the form of a transport problem according to the time criterion [14].

There exists m of starting points (SP) A_1, \dots, A_m with margin a_1, \dots, a_m and n of destination points (DP) B_1, \dots, B_m with requests b_1, \dots, b_m the sum of margins equals the sum of requests:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j.$$

The times of transportation t_{ij} from each SP A_i to each DP B_j are given, it is assumed that they do not depend on the amount of the transported cargo.

It is required to choose transportation (x_{ij}) in such a way that the balance conditions are met

$$\begin{cases} \sum_{j=0}^n x_{ij} = a_i \quad (i = 1, \dots, m), \\ \sum_{i=0}^m x_{ij} = b_j \quad (j = 1, \dots, m), \end{cases}$$

and in addition, the end time of all transportations T turned at a minimum. Thus, it is necessary to find a transport plan (x_{ij}) for which the time T turns into a minimum:

$$T = \max_{x_{ij} > 0} t_{ij} = \min.$$

The described problem can be used to select the optimal route for computing problems at time t^0 as follows. Computing nodes will be software A_1, \dots, A_m , and their free computing resource at the moment t^0 will be a stock in terms of transport problems. Moreover, each task will be a DP with a certain need for computing resources. This introduces an additional condition for transportation – each DP must be served by single software.

In order to support the heterogeneity of both computational tasks and software and hardware, computational tasks are considered as atomic – that is, indivisible between CNs. If a computation task is not atomic, it must be represented as a set of sequentially (or in parallel, depending on the nature of the task) of atomic applications running.

The condition on the atomicity of tasks introduces into the formulation of the transport problem the above-mentioned restriction that each DP must be serviced by a single software; in what follows, this type of transport problem will be called a transport problem by the time criterion with atomic needs.

The atomicity condition makes it possible not to consider the combination of servicing the DP by several softwares, which significantly reduces the complexity of the solution algorithm in comparison with the classical solution of the transport problem by the time criterion.

The transportation time (in terms of the model, the processing time of the task) t_{ij} is calculated as the sum of the task B_j delivery time forecast to the node A_i , the forecast of the calculation time, and the forecast of the delivery time of the received data to the destination.

If at time t^0 the computational task B_j is being processed at the node A_i , t_{ij} will express the remaining processing time + the time of data delivery to the end point.

For other nodes, the processing time for this task will include the cost of transferring the calculation from the current node to another.

If the CN A_i cannot process the task B_j , for example, does not have the necessary software, then we assume that the processing time t_{ij} is equal to infinity.

Least-time algorithm for atomic claims. The problem of choosing the most efficient computation route at time t^0 can be represented as a transport problem according to time criterion with atomic needs described above. Taking into account the condition that computational tasks are atomic, that is one task can be processed only on one node (in practice, this can be achieved by preliminary partitioning of complex calculations into a sequence of atomic tasks), the solution of a transport problem by the time criterion with atomic needs degenerate. The developed least-time algorithm for atomic claims (LTfAC) for the transport problem according to the time criterion with atomic needs can be represented as a diagram in fig. 3.

The algorithm works as follows. It is necessary to select a pair of DP and software with the smallest t_{ij} , provided that the software has all the necessary resources to service the DP, and then adjust the stock for software by the amount of DP and repeat the selection until all DP are served, or until no software will be able to serve the remaining DP. In this case, tasks for which a suitable computational node is determined are sent for computation, and the remaining tasks wait for the next iteration of the algorithm.

The efficiency of the developed algorithm was investigated on the HDIPS simulation model [15]. Fig. 4 shows a graph of the average computation time of tasks in the HDIPS dependence on its structure when using the LTfAC algorithm as an algorithm for computing resources distribution in HDIPS. For clarity, the axes of CN number and the number of tasks sources on the graph are inverted.

The developed algorithm provides the minimum average time for tasks completion and this time is fairly stable relative to the number of CNs. Thus, the developed algorithm is recommended to be used if the advantage from reducing the time of tasks calculations exceeds the cost of additional HDIPS computational load, thus the J^2 optimality criterion of the configuration is selected.

Conclusion. The following possible criteria for the optimality of the HDIPS software and hardware configuration were identified and formalized:

- minimizing the average value of the CN utilization factor;
- minimizing the total time for problem solving in the system;
- minimizing the probability of problem return with a CN due to a lack of computing resources.

To meet the selected optimality criterion it is proposed to analyze the HDIPS software and hardware configuration using its formal model applying queuing theory methods. HDIPS can be represented as a queuing network, where each CN and route control unit is a mass service system. In terms of the proposed formal model, a function is defined which expresses the probability of a task transition from a router to a computational node. The example of the presentation of the criterion “minimizing

the total time of problem solving in the system” in the formal model using the apparatus of the queuing theory is given.

The problem of computing resource allocation in HDIPS is presented as a transport problem according to the time criterion with atomic needs. The atomicity condi-

tion makes it possible not to consider the combination of serving one destination (in terms of the transport problem) by several starting points, which significantly reduces the complexity of the solution algorithm in comparison with the classical solution of the transport problem by the time criterion.

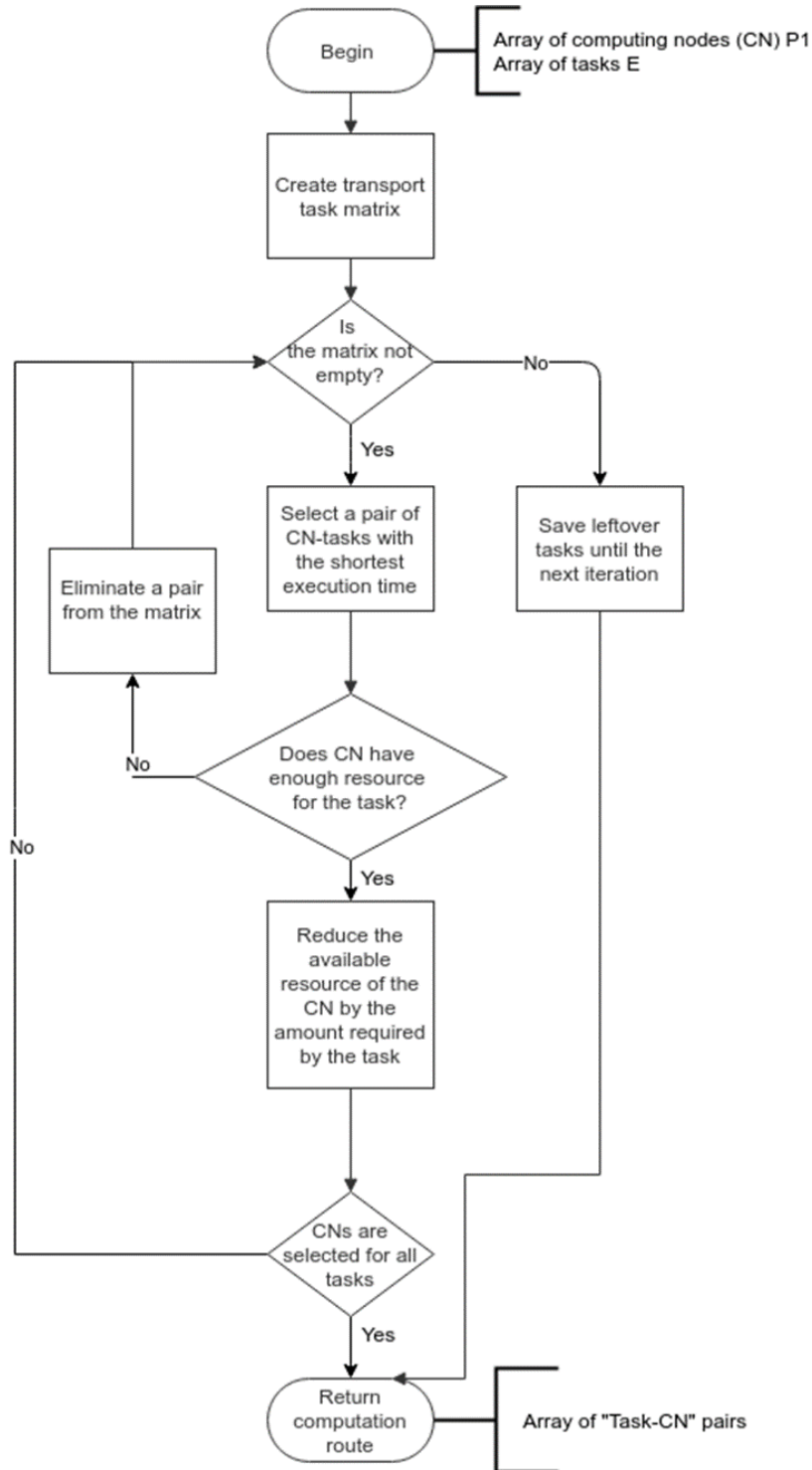


Fig. 3. Block diagram of the resource allocation algorithm via solving a transport problem by the time criterion with atomic needs using the LTfAC method

Рис. 3. Блок-схема алгоритма распределения ресурсов решением транспортной задачи по критерию времени с атомарными потребностями методом НВдАЗ

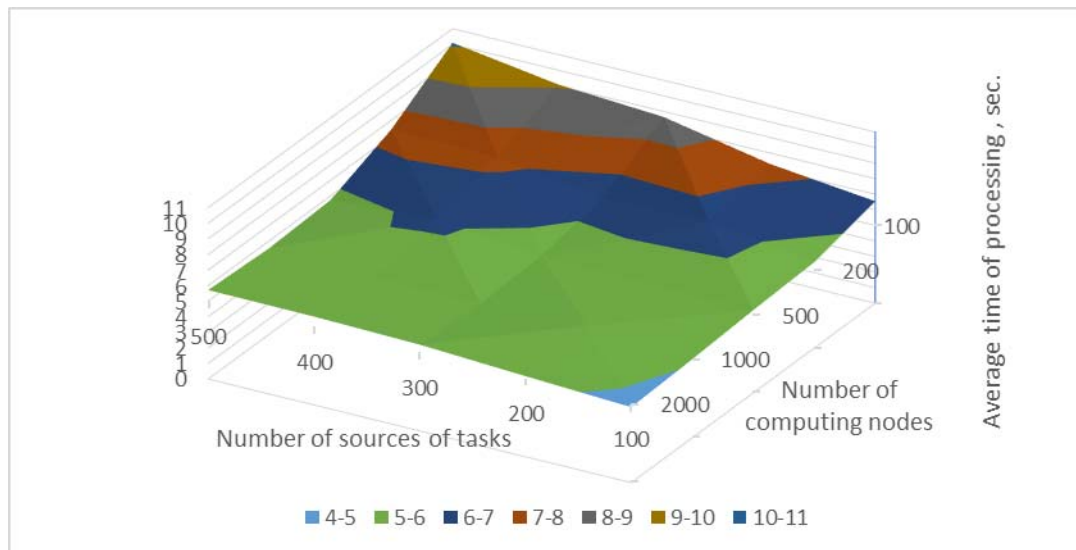


Fig. 4. Time of tasks processing when using the LTfAC algorithm for resource allocation

Рис. 4. Время обработки задач при использовании алгоритма НВДАЗ для распределения ресурсов

Taking into account the condition of atomicity, the least-time algorithm for atomic claims was developed. In comparison with other investigated algorithms, the developed algorithm provides the minimum average time for executing tasks and this time is fairly stable relative to the number of CNs. Thus, the developed algorithm is recommended to be used if the advantage from reducing the time for calculating the tasks exceeds the cost of the additional computational load of the HDIPS, that is, the criterion of optimality of the configuration “minimizing the total time for problem solving in the system” is selected.

References

1. Antamoshkin O. A., Kilochitskaya T. R., Ontuzheva G. A., Stupina A. A., Tynchenko V. S. Multicriterion problem of allocation of resources in the heterogeneous distributed information processing systems. *Journal of Physics: Conference Series*. 2018, Vol. 1015, P. 32162. Doi: 10.1088/1742-6596/1015/3/032162.
2. Antamoshkin O. A. [Multi-agent automation system for monitoring, forecasting and control in emergency situations]. *Mezhdunarodna nauchna shkola “Paradigma”* [Paradigma International Scientific School]. Varna, 2015, P. 18–28 (In Russ.).
3. Glazunov V. V., Kurochkin M. A., Popov S. G. [Method for evaluating message transmission routes in telematic networks of vehicles based on the logical-probabilistic method]. *Intellektual'nye tekhnologii na transporte*, 2015. Vol 1 (In Russ.). Available at: <https://cyberleninka.ru/article/n/metod-otsenki-marshrutov-peredachi-soobscheniy-v-telematicheskikh-setyah-transportnyh-sredstv-na-osnove-logiko-veroyatnostnogo-metoda> (accessed: 25.10.2020).
4. Bigham J., Du L. Cooperative negotiation in a multi-agent system for real-time load balancing of a mobile cellular network *ACM*, 2003. P. 568–575. Doi: 10.1145/860575.860666.
5. Kantamneni A., Brown L. E., Parker G., Weaver W. W. Survey of multi-agent systems for microgrid control. *Engineering applications of artificial intelligence*. 2015, Vol. 45, P. 192–203. Doi: 10.1016/j.engappai.2015.07.005.
6. Khritankov A. S. [Modeli i algoritmy raspredeleniya nagruzki]. *Informatsionnye tekhnologii i vychislitel'nye sistemy*. 2009, Vol. 2, P. 65–80 (In Russ.).
7. Skobelev P. O. [Intelligent resource management systems in real time: development principles, experience of industrial implementations and development prospects]. *Prilozhenie k teoreticheskomu i prikladnomu nauchno-tekhnicheskomu zhurnalu “Informatsionnye tekhnologii”*. 2013, No. 1, P. 1–32 (In Russ.).
8. Dmitriev V. N., Sorokin A. A., Kuok Ch. T. [Improving the efficiency of traffic management in heterogeneous data transmission systems under conditions of uncertainty]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika*. 2015, No. 3, P. 66–77 (In Russ.).
9. Krutolapov A. S. [Ensuring the quality of service in information exchange networks]. *Vestnik Voronezhskogo instituta GPS MChS Rossii*. 2013, No. 1, P. 18–22 (In Russ.).
10. Kammoun H. M., Kallel I., Casillas J., Abraham A., Alimi A. M. Adapt-Traf: An adaptive multiagent road traffic management system based on hybrid antihierarchical fuzzy model. *Transportation Research Part C: Emerging Technologies*. 2014, No. 42, P. 147–167. Doi.org: 10.1016/j.trc.2014.03.003.
11. *GOST 15971–90. Sistemy obrabotki informatsii. Terminy i opredeleniya* [State Standard 15971–90. Information processing systems. Terms and Definitions]. Moscow, Standartinform Publ., 1991, 12 p.
12. Ontuzheva G. A. [Methods for optimizing the distribution of resources of a geographically distributed multi-level computer network]. *Mezhdunarodna nauchna*

shkola "Paradigma" [Paradigma International Scientific School]. Varna, 2015, P. 185–190 (In Russ.).

13. Zhozhikashvili V. A., Vishnevskiy V. M. [Queuing networks: Theory and application to computer networks]. *Seti massovogo obsluzhivaniya: Teoriya i primeneniye k setyam EVM*. Moscow, Radio i svyaz' Publ., 1988, 191 p.

14. Hammer P. L. Timeminimizing transportation problems. *Naval Research Logistics Quarterly*. 1969, No. 3 (16), P. 345–357. Doi:10.1002/nav.3800160307.

15. Ontuzheva G. A., Bruchanova E. R., Rudov I. N., Pikov N. O., Antamoshkin O. A. Simulation modelling of the heterogeneous distributed information processing systems. *In IOP Conference Series: Materials Science and Engineering*. 2018, Vol. 450, No. 5, P. 05. Doi: 10.1088/1757-899X/450/5/052018.

Библиографические ссылки

1. Multicriterion problem of allocation of resources in the heterogeneous distributed information processing systems / O. A. Antamoshkin, T. R. Kilochitskaya, G. A. Ontuzheva et al. // *Journal of Physics: Conference Series*. 2018. Vol. 1015. P. 32162.

2. Антамошкин О. А. Мультиагентная система автоматизации мониторинга, прогнозирования и управления в чрезвычайных ситуациях // Международная научная школа «Парадигма». Лято-2015, 2015.

3. Глазунов В. В., Курочкин М. А., Попов С. Г. Метод оценки маршрутов передачи сообщений в телематических сетях транспортных средств на основе логико-вероятностного метода [Электронный ресурс] // *Интеллектуальные технологии на транспорте*. 2015. № 1. URL: <https://cyberleninka.ru/article/n/metod-otsenki-marshrutov-peredachi-soobscheniy-v-telematicheskikh-setyah-transportnyh-sredstv-na-osnove-logiko-veroyatnostnogo-metoda> (дата обращения: 25.10.2020).

4. Bigam J., Du L. Cooperative negotiation in a multi-agent system for real-time load balancing of a mobile cellular network ACM, 2003. P. 568–575.

5. Kantamneni A., Brown L. E., Parker G., Weaver W. W. Survey of multi-agent systems for microgrid control // *Engineering applications of artificial intelligence*. 2015. Vol. 45. P. 192–203.

6. Хританков А. С. Модели и алгоритмы распределения нагрузки // *Информационные технологии и вычислительные системы*. 2009. № 2. С. 65–80.

7. Скобелев П. О. Интеллектуальные системы управления ресурсами в реальном времени: принципы разработки, опыт промышленных внедрений и перспективы развития // Приложение к теоретическому и прикладному научно-техническому журналу «Информационные технологии». 2013. № 1. С. 1–32.

8. Повышение эффективности управления трафиком в гетерогенных системах передачи данных в условиях неопределенности / В. Н. Дмитриев, А. А. Соколин, Ч. Т. Куок и др. // *Вестник Астраханского гос. техн. ун-та. Серия: Управление, вычислительная техника и информатика*. 2015. № 3. С. 66–77.

9. Крутолапов А. С. Обеспечение качества обслуживания в сетях информационного обмена // *Вестник Воронежского ин-та ГПС МЧС России*. 2013. № 1. С. 18–22.

10. Adapt-Traf: An adaptive multiagent road traffic management system based on hybrid ant-hierarchical fuzzy model / Н. М. Kammoun, I. Kallel, J. Casillas et al. // *Transportation Research Part C: Emerging Technologies*. 2014. No. 42. P. 147–167.

11. ГОСТ 15971–90. Системы обработки информации. Термины и определения. Взамен ГОСТ 15971–84. Дата введения в действие: 01.01.1992. Статус документа – действующий. Дата издания: 11.01.1991. Дата последнего изменения: 19.04.2010. М. : Изд-во стандартов, 1991. 12 с

12. Онтужева Г. А. Методы оптимизации распределения ресурсов территориально распределенной многоуровневой вычислительной сети // Международная научная школа «Парадигма». Лято-2015, 2015. С. 185–190.

13. Жожикашвили В. А., Вишневский В. М. Сети массового обслуживания: Теория и применение к сетям ЭВМ. М. : Радио и связь, 1988.

14. Hammer P. L. Timeminimizing transportation problems // *Naval Research Logistics Quarterly*. 1969. № 3 (16). P. 345–357.

15. Simulation modelling of the heterogeneous distributed information processing systems / G. A. Ontuzheva, E. R. Bruchanova, I. N. Rudov et al. // *In IOP Conference Series: Materials Science and Engineering*. 2018. Vol. 450, No. 5. P. 05.

© Ontuzheva G. A., 2020

Ontuzheva Galina Aleksandrovna – Assistant of the Department of Information Technologies in Creative and Cultural Industries; Siberian Federal University. E-mail: gontuzheva@sfu-kras.ru.

Онтужева Галина Александровна – ассистент кафедры информационных технологий в креативных и культурных индустриях; Сибирский федеральный университет. E-mail: gontuzheva@sfu-kras.ru.
