

Формирование комплектов эффективных алгоритмов распределения вычислительных ресурсов в гетерогенных динамичных вычислительных средах на основе онтологии

А. Б. Клименко ¹ ✉, Э. М. Алиева ¹, А. Е. Сальников ¹

¹ Институт информационных наук и технологий безопасности
Российского государственного гуманитарного университета,
ул. Академика Янгела, д. 25, к. 2, г. Москва 117534, Российская Федерация

✉ e-mail: anna_klimenko@mail.ru

Резюме

Цель исследования. Цель данного исследования – разработка структуры онтологии как основы базы данных/базы знаний для выбора эффективных метаэвристических алгоритмов решения задачи распределения нагрузки в гетерогенных распределенных динамичных вычислительных средах с учетом накладных расходов на передачу данных по сети.

Методы. Основными научными методами, применяемыми в рамках данного исследования, являются анализ предметной области, методы построения предметных онтологий, численные методы оптимизации и компьютерное моделирование. Поскольку в литературе не представлены модели планирования выделения ресурсов, которые бы учитывали географическую распределенность, наличие промежуточных маршрутов передачи данных, динамику топологий и нагрузки, а также гетерогенность системы в аспекте критериев оценивания качества распределения нагрузки, в данной статье предложена новая модель, учитывающая перечисленные особенности. Трудоемкость решения задачи планирования становится одним из варьируемых параметров, который при этом оказывает значительное влияние на результат планирования: с уменьшением трудоемкости вычислений, соответственно ухудшается результат. Поэтому в качестве метода решения предлагается жадная стратегия: из подлежащих рассмотрению методов оптимизации выбрать такой наименее трудоемкий, который бы позволял получить наилучший результат за выделенное время. Тестовые запуски алгоритмов имитации отжига демонстрируют различную эффективность на различных исходных условиях задачи, следовательно, целесообразно для выделенных классов задач выбрать эффективные в смысле качества решения и трудоемкости алгоритмы.

Результаты. Результатом исследования является структура онтологии эффективных алгоритмов. Также результатами являются вошедшие в состав онтологии экземпляры алгоритмов имитации отжига и задач, связанные отношением «эффективности».

Заключение. В данной статье предложены структура онтологии эффективных алгоритмов оптимизации и подход к решению задачи распределения вычислительной нагрузки с учетом трудоемкости процедуры распределения посредством «жадного» выбора наиболее эффективных алгоритмов оптимизации.

Ключевые слова: распределенные вычисления; онтология; методы оптимизации; метаэвристики; распределение нагрузки.

Конфликт интересов: Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Клименко А. Б., Алиева Э. М., Сальников А. Е. Формирование комплектов эффективных алгоритмов распределения вычислительных ресурсов в гетерогенных динамических вычислительных средах на основе онтологии // Известия Юго-Западного государственного университета. 2023; 27(4): 25-43. <https://doi.org/10.21869/2223-1560-2023-27-4-25-43>.

Поступила в редакцию 05.10.2023

Подписана в печать 17.11.2023

Опубликована 21.12.2023

Efficient Algorithm Set Forming for the Computing Resources Distribution in Heterogeneous Dynamic Computational Environments Based on the Ontology Usage

Anna B. Klimenko ¹ ✉, Elvira M. Alieva ¹, Andrey Y. Salnikov ¹

¹ Institute of IT and Security Technologies of Russian State University for the Humanities
25, building 2 Academician Yangel str., Moscow 117534, Russian Federation

✉ e-mail: anna_klimenko@mail.ru

Abstract

Purpose of research. The purpose of this research is to develop an ontology structure as the basis of a database/knowledge base for selecting effective metaheuristic algorithms for solving the problem of load distribution in heterogeneous distributed dynamic computing environments, taking into account the overhead of data transmission over the network.

Methods. The main scientific methods used in this study are domain analysis, methods for constructing subject ontologies, numerical optimization methods and computer modeling.

Since the literature does not present resource allocation planning models that would take into account geographic distribution, the presence of intermediate data transmission routes, the dynamics of topologies and load, as well as system heterogeneity in terms of criteria for assessing the quality of load distribution, this article proposes a new model that takes into account these features. The complexity of solving a planning problem becomes one of the variable parameters, which has a significant impact on the planning result: with a decrease in the complexity of calculations, the result deteriorates accordingly. Therefore, a greedy strategy is proposed as a solution method: from the optimization methods to be considered, select the least labor-intensive one that would allow obtaining the best result in the allotted time. Test runs of simulated annealing algorithms demonstrate different effectiveness under different initial conditions of the problem; therefore, it is advisable for selected classes of problems to choose algorithms that are effective in terms of solution quality and labor intensity.

Results. The result of the study is the structure of the ontology of effective algorithms. Also, the results are instances of simulated annealing algorithms and tasks included in the ontology, related by the “efficiency” relation.

Conclusion. This article proposes the structure of an ontology of effective optimization algorithms and an approach to solving the problem of distributing the computational load, taking into account the complexity of the distribution procedure through the “greedy” selection of the most effective optimization algorithms.

Keywords: *distributed computing; ontology; optimization methods; metaheuristics; workload distribution.*

Conflict of interest. *The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.*

For citation: Klimenko A. B., Alieva E. M., Salnikov A. Y. Efficient Algorithm Set Forming for the Computing Resources Distribution in Heterogeneous Dynamic Computational Environments Based on the Ontology Usage. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2023; 27(4): 25-43 (In Russ.). <https://doi.org/10.21869/2223-1560-2023-27-4-25-43>.

Received 05.10.2023

Accepted 17.11.2023

Published 21.12.2023

Введение

В настоящее время одним из трендов в области распределенных вычислений является частичный отказ от «чистых» облачных вычислений и их расширение путем перехода к концепциям «туманных» и «краевых» вычислений. Это объясняется недостаточностью «облачной» концепции в ситуации, когда пользовательские приложения имеют достаточно жесткие ограничения на время выполнения пользовательских операций, а географическая удаленность датацентра и наличие транзитных (и, как правило, гетерогенных) участков сети делают проблематичным гарантированный временной интервал обслуживания [1]. Последнее негативно сказывается на качестве сервиса (Quality of Service) и качестве опыта использования (Quality of Experience), что, в свою очередь, ведет к финансовым потерям [2].

Однако использование туманных и краевых вычислений имеет существенные и весомые отличия от облачных, а именно: высокая степень гетерогенности вычислительных устройств, высокая степень гетерогенности каналов

связи, а также относительно высокая динамика как сетевых топологий, так и номенклатуры пользовательских задач. Кроме того, вычислительные устройства в туманной среде, как правило, менее производительны, чем устройства в датацентрах и распределены географически [3].

Географическая распределенность устройств в составе туманного слоя является причиной возникновения дополнительной нагрузки на устройства, в тех случаях, когда они используются для передачи данных соседним узлам.

Высокая динамика туманной вычислительной среды актуализирует вопрос о частой смене конфигурации системы распределенных вычислений, то есть, закрепления задач за узлами (в случае появления новых задач), а также перераспределения задач (например, в том случае, когда пользователь изменяет географическую позицию). В качестве примера можно привести ситуацию пользовательского приложения, работающего в едущем поезде, который следует мимо вышек 5G сотовой связи, либо вычислений, которые могут про-

изводиться в группах низколетящих спутников с ретрансляцией.

Гетерогенность вычислительной среды, в свою очередь, ставит вопрос о необходимости рассмотрения задачи распределения вычислительных ресурсов не только как многокритериальной задачи оптимизации, но и как задачи, где, возможно, отдельно взятые вычислительные узлы, помимо общего критерия оптимизации, будут иметь свои собственные.

Таким образом, постепенное формирование и интенсификация использования вычислительных сред, обладающих ярко выраженными свойствами гетерогенности (в том числе, гетерогенности критериев эффективности распределения ресурсов), динамичности и географической распределенности, является новым этапом в развитии области распределенных вычислений, делая необходимым разработку комплекса методов управления ресурсами с учетом перечисленных особенностей.

Данная статья является результатом исследований, проводимых в области моделей и методов распределения вычислительных ресурсов. На основе аналитического обзора литературы сделаны выводы о моделях и методах управления ресурсами гетерогенных динамичных распределенных вычислительных сред как о частных реализациях, и представлена модель общей задачи управления вычислительными ресурсами. На основе анализа сформулированной задачи сделан вывод о том, что блок

подбора процедуры смены конфигурации вычислительной системы может быть реализован посредством применения технологий искусственного интеллекта, где самым простым подходом является использование продукционных правил. При этом принятие решения о выборе того или иного алгоритма оптимизации осуществляется путем выбора вариантов из сформированного портфолио эффективных алгоритмов. Последнее реализуется на основе онтологической модели предметной области решения задач распределения нагрузки.

Цель данной статьи – разработка структуры онтологии предметной области выбора эффективных алгоритмов для распределения вычислительных ресурсов в распределенных гетерогенных динамичных вычислительных средах.

Материалы и методы

Формализация задач распределения нагрузки

От начала своего появления в 2012 г. и до нынешних дней задачи управления ресурсами в туманных вычислительных средах не утратили актуальности. В рамках распределенных вычислений туманные вычисления воплощают наиболее сложный случай с точки зрения управления ресурсами по сравнению, например, с GRID-системами [4] – потому, что к свойствам гетерогенности и распределенности добавляется свойственное облачным сервисам критерии экономической выгоды, непосредственно с «облаком» - в силу гетерогенности коммуни-

кационной среды, наличия транзитных узлов и, в целом, более низкой пропускной способности. Если сравнивать с мультимпьютером [5], то системы туманных вычислений также существенно усложняются относительно высокой степени географической распределенности, а также потенциальной динамикой топологии. Эти особенности – гетерогенность структурная и критериев, динамика нагрузки и топологий, географическая распределенность – фактически делают ранее разработанные модели и методы малоэффективными по следующим причинам:

– если речь идет о составлении расписаний или распределении нагрузки, то, для статичных систем расписание составляется однократно, либо, если подразумевается реконфигурирование системы, то как реакция на изменения среды, без учета, собственно, сложности процедуры;

– также не учитывается специфика коммуникационной среды, а именно – наличие транзитных участков сети, объединенных устройствами, которые получают дополнительную нагрузку и, соответственно, тратят определенные ресурсы;

– как правило, гетерогенный характер сети учитывается в качестве ограничений, за исключением некоторых работ, где предложено использовать аддитивную свертку разнообразных критериев.

Среди фундаментальных работ, посвященных распределенным вычислениям, но не учитывающих перечисленное, можно выделить следующие [6-10].

Анализ направлений в рамках управления ресурсами туманных вычислений показал, что в основном выбраны следующие направления [11]:

- Обеспечение ресурсами.
- Размещение приложений.
- Разгрузка задач.
- Планирование.
- Балансировка нагрузки.
- Выделение ресурсов.

При этом рассмотрение работы [11] позволяет выделить следующие критерии и методы решения задач распределения нагрузки:

- Время завершения решения комплекса задач.
- Стоимость комплекса вычислительных устройств.
- Энергопотребление.
- Время реакции системы на внешнее воздействие.
- Балансировка нагрузки.

Имеются как однокритериальные постановки задач, так и многокритериальные, а именно: в [12] рассмотрена постановка задачи двухкритериальной оптимизации, где критерии – время выполнения задач и стоимость.

В работе [13] производится оптимизация распределения нагрузки по критериям времени выполнения и энергопотребления.

В [14] рассмотрена постановка задачи с оптимизацией по критериям времени выполнения комплекса задач и балансировки нагрузки.

Авторы [15] также рассматривают задачу оптимизации времени выполнения и потребляемой задачами памяти.

Авторы [16] представили постановку задачи и метод решения для трех критериев: время реакции системы, время выполнения задач и стоимости.

Работы [17-20] посвящены также одно- и двухкритериальной оптимизации распределения вычислительной нагрузки.

Следует отметить, что в перечисленных работах также представлены методы решения, как правило, являющиеся реализацией того или иного метаэвристического метода, адаптированного для наиболее эффективного решения поставленной задачи.

Использовались генетические алгоритмы, алгоритм «кукушкин поиск», оптимизация «серых волков», муравьиных колоний, имитации жизни пчел и др.

Рассмотренные задачи были решены без учета, собственно, времени на решение задачи, а также без учета того времени, которое потребуется для перемещения программного кода задач и данных с узла на узел, в то время как время решения задачи в условиях динамики системы может приводить к тому, что будут нарушены ограничения на время выполнения пользовательских задач.

Таким образом, по результатам аналитического обзора могут быть сделаны следующие выводы:

– представленные модели задач управления ресурсами и методы их решения не отражают таких особенностей современных систем распределенных вычислений, как: распределенность, гетерогенность и динамика;

– метод изменения конфигурации вычислительной системы влияет как на эффективность результирующего распределения нагрузки, так и на время, потребляемое на реконфигурацию, и должен быть учтен при выборе метода распределения ресурсов.

Общая постановка задачи управления ресурсами системы распределенных вычислений и формирование комплектов эффективных алгоритмов распределения нагрузки

Для распределения нагрузки по фрагменту географически распределенной сети используются следующие входные данные.

- Граф задач, предназначенных к решению: $G_1 = \{ \langle g_i, r_i \rangle, R \}$, где g_i – вычислительная сложность задачи; r_i – требования к ресурсам устройства, на котором будет происходить размещение, включая: требования к объему памяти, требования к пропускной способности канала, к производительности и т.д. Данный граф – ациклический и направленный, где ребра R взвешены объемом передаваемых данных между задачами.

- Граф сети представляется произвольным направленным мультиграфом, где вершины взвешены характеристиками узлов сети (производительность, объем памяти, энергопотребление и т.д.), ребра взвешены скоростями передачи данных по каналам связи. То есть: $G_2 = \{ M, C \}$, $M = \{ m_j \}$ – ресурсы, которыми располагает узел, $C = \{ c_j \}$ – каналы связи.

- Имеются общие критерии оценивания качества распределения задач $S_0 = \{s_k\}$.

- Гетерогенность сети и специфика используемых устройств продуцирует индивидуальные критерии качества распределения, специфичные отдельным узлам, и составляет множество: $P_0 = \{p_l\}$.

- Имеются общие ограничения $constr = \{constr_k\}$.

- Процедура управления характеризуется параметрами $\langle g_r, r_r, t_r \rangle$, где g_r – вычислительная сложность задачи перепланирования; r_r – требования к ресурсам узла, где выполняется расчет нового закрепления № t_r – время выполнения перепланирования (перемещение данных).

- Остатки ресурсов транзитных узлов определяется как:

$$\forall i, k, L, m_{i_ост} = m_i - \sum_{k=1}^L r_k, \text{ где } L - \text{ количество задач, для которых узел } i \text{ включен в маршрут передачи данных.}$$

- Таким образом, управление системой РВ будет заключаться в решении следующей задачи: необходимо для графов G_1 и G_2 найти такие закрепление задач за устройствами и $\langle g_r, t_r \rangle$, чтобы при имеющихся ограничениях: $r_i \leq m_j$, $constr_j$, $r_r \leq m_j$ обеспечить $S_0 \rightarrow \max$, $P_0 \rightarrow \max$, $m_{i_ост} \rightarrow \max, \forall i \in \{Ro_\alpha\}$, где $\{Ro_\alpha\}$ – множество маршрутов.

При такой постановке задачи решение ее классическими методами оптимизации не представляется возможным, потому что, если вести речь об итерационных методах, то результат решения

задачи распределения нагрузки улучшается с увеличением количества итераций (поколений/популяции). При этом увеличивается время собственно получения такого решения. Также, несмотря на имеющиеся оценки времени сходимости тех или иных алгоритмов, при уменьшении времени их работы заранее неизвестно, насколько хорошим или плохим будет полученное решение.

Поэтому предлагается в качестве стратегии выбрать «жадный» способ выбора алгоритмов решения задачи распределения нагрузки, а именно: в зависимости от признаков задачи осуществлять выбор того метаэвристического метода с такими параметрами, который бы позволял получить наилучшее решение при наименьшем количестве итераций. Такой подход требует проведения экспериментального исследования наборов метаэвристик, опираясь на множество выбранных признаков, и формирования комплектов эффективных алгоритмов для заданной задачи. Рациональным подходом к решению такой задачи представляется анализ предметной области, разработка онтологии и затем, на базе полученной онтологии, формирования отношений между экземплярами задач и экземплярами метаэвристических методов их решения.

Эффективность алгоритмов
в решении поставленной задачи

Различная эффективность алгоритмов для задач распределения нагрузки с

различными параметрами подтверждается экспериментально, например: производится распределение нагрузки по узлам сети топологии «решетка» с увеличением количества узлов по свертке индивидуальных показателей вероятности безотказной работы (ВБР) узлов, взятой в произвольный момент времени $t=100$ ч., и производится сравнение ре-

зультатов работы простого генетического алгоритма, случайного поиска (метод Монте-Карло), метода имитации отжига с температурной схемой тушения.

Как видно из табл. 1, получены значения свертки одного порядка, при этом время, затраченное на поиск решений, существенным образом различается, что показано в табл. 2.

Таблица 1. Сравнение результатов распределения задач по узлам

Table 1. The comparison of tasks distribution results

Кол-во узлов сети / Number of network nodes	Кол-во задач / Number of tasks	Результат работы ГА / Result of GA work	Результат работы СП / Result of JV work	Результат работы ИО / Result of IO work
100	10	0.98552	0.97349	0.96091

Таблица 2. Сравнение времени работы алгоритмов

Table 2. The comparison of algorithms time consumption

Кол-во узлов сети / Number of network nodes	Кол-во экспериментов / Number of experiments	Время работы ГА (с) / GA operating time (s)	Время работы СП (с) / SP operating time (s)	Время работы ИО (с) / IO operating time (s)
100	10	79.27919	68.62277	1.27247

В качестве примера также рассмотрим результаты тестов с различными типами алгоритмов имитации отжига.

Еще один эксперимент, демонстрирующий различия в эффективности применения метаэвристик на различных входных данных, описан ниже. Исследованы алгоритмы имитации отжига со следующими параметрами:

1. Закон изменения температуры:

– геометрическое понижение температуры:

$$T_{\text{new}} = \alpha * T_n, \text{ где } \alpha \in (0,1);$$

– экспоненциальное понижение температуры:

$$T_{\text{new}} = T_n * e^{-B/n}, \text{ где } B = \frac{-\ln\left(\frac{T_{\text{min}}}{T_n}\right)}{k},$$

где $k > 0$;

– логарифмическое понижение температуры:

$$T_{\text{new}} = T_n - k * \log(n+1);$$

– квадратичное понижение температуры:

$$T_{\text{new}} = T_n - k * n^2.$$

2. Формулы вероятностного перехода:

– линейная функция Больцмана:

$$P = 1 - \frac{\Delta E}{T_n};$$

– экспоненциальная функция Больцмана:

$$P = e^{\frac{-\Delta E}{T_n}}.$$

Для первого эксперимента было взято пять вычислительных узлов про-

извольной производительности и 50 задач с небольшой амплитудой изменения трудоемкостей [40;50].

В заданных условиях независимо от способа изменения температуры и количества итераций лучшие результаты показал алгоритм, использующий экспоненциальную функцию Больцмана (табл. 3, рис. 1).

Таблица 3. Средние значения ЦФ эксперимента 1

Table 3. Mean values of the objective function in simulation 1

Вер. Итер.	Геом. пониж.		Экспон. пониж.		Лог. пониж.		Квадр. пониж.	
	ЛФБ	ЭФБ	ЛФБ	ЭФБ	ЛФБ	ЭФБ	ЛФБ	ЭФБ
100	1,945	1,558	1,558	1,558	2,411	1,558	1,918	1,558
1000	1,671	1,558	1,558	1,558	1,726	1,558	1,849	1,558
5000	1,558	1,558	1,558	1,558	1,558	1,558	1,766	1,558

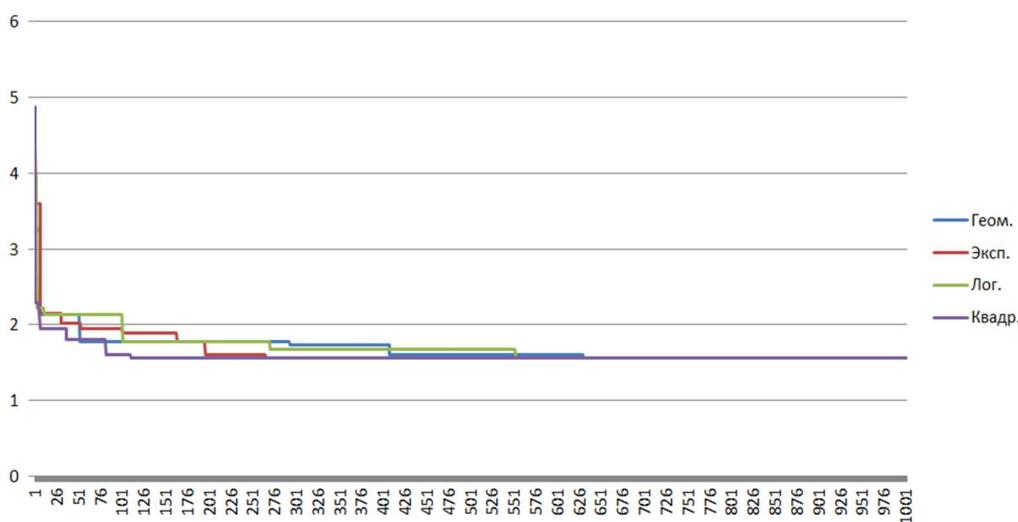


Рис.1. Пример результатов эксперимента с применением линейной функции Больцмана, 1000 итераций

Fig.1. The example of simulation results with the usage of Boltzmann function, 1000 iterations

В ходе следующего эксперимента была рассмотрена система, в состав которой входило 20 вычислительных узлов и 100 задач с двумя вариантами интервалов значений трудоемкостей: [40,50] и

[40,100]. Количество тестов для каждого эксперимента – 100. Полученные средние показатели значений ЦФ приведены в табл. 4.

Таблица 4. Результаты эксперимента с различными интервалами значений трудоемкостей задач**Table 4.** Simulation results with the various computational complexities of tasks

	Темп.	Геом. пониж.		Экспон. пониж.		Лог. пониж.		Квадр. пониж.		
		Вер. Итер.	ЛФБ	ЭФБ	ЛФБ	ЭФБ	ЛФБ	ЭФБ	ЛФБ	ЭФБ
Мал.	100		8,93	6,44	8,6	8,2	9,33	9,33	9,27	9,07
	1000		6,47	6,61	6,97	7,87	8,8	7,13	8,65	6,5
	5000		6,23	5,87	6,97	6,57	6,71	6,57	6,14	6,22
Бол.	100		8,37	9,38	9,77	8,7	9,52	8,74	10	9,86
	1000		7,06	7,7	8,20	8,14	9,79	8,4	9,37	8,14
	5000		7,52	7,65	7,84	7,23	8,93	7,8	9,52	7,41

То есть, по результатам экспериментов на одном и том же временном интервале функционирования процедуры поиска возможно улучшение найденного решения до 30%.

Пример, иллюстрирующий изменение значений ИО с экспоненциальной

вероятностной формулой Больцмана, показан на рис. 2.

Видно, что в зависимости от изменений исходных условий задачи, менялась и эффективность применения тех или иных типов алгоритмов ИО в условиях ограничения на время.

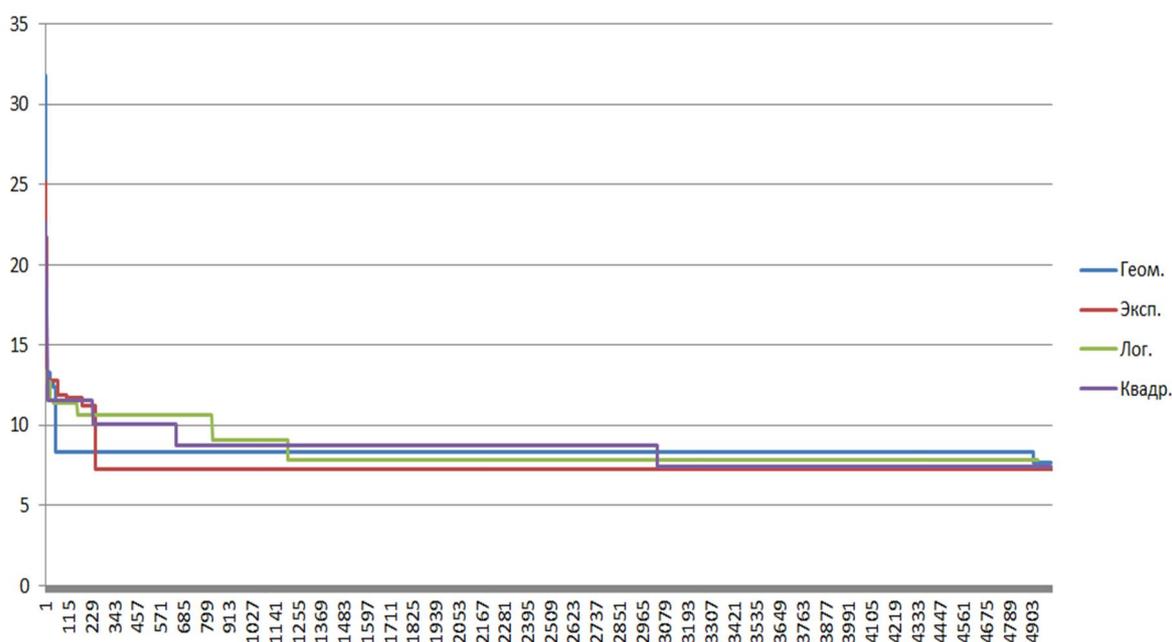


Рис. 2. Изменение значений целевой функции в зависимости от количества итераций ИО, с использованием экспоненциальной вероятностной формулы, 5000 итераций

Fig. 2. The objective function values change depending on the SA iterations number, with the exponential law and 5000 iterations usage

То есть, в условиях малой вариативности трудоемкостей задач, на малом количестве итераций (100) наиболее эффективной комбинацией параметров является понижение температуры по геометрическому закону при использовании экспоненциальной функции Больцмана, в то время как при большой вариативности трудоемкостей задач наилучший результат на аналогичном количестве итераций позволило получить использование линейной функции Больцмана в совокупности с геометрическим законом понижения температуры.

Таким образом, располагая результатами экспериментальных исследований эффективности различных алгоритмов, применяемых в данной области, имеется возможность, на основе онтологии, и в рамках этой онтологии, на основе отношений между экземплярами задач и методов их решения, создание комплектов эффективных алгоритмов, применимых для решения задачи распределения нагрузки в гетерогенной, динамичной и распределенной среде в соответствии с формулировкой задачи и в рамках «жадной» стратегии.

Результаты и их обсуждение

Онтология предметной области решения задач распределения нагрузки

Основными классами онтологии предметной области решения задач распределения нагрузки является класс «Алгоритмы» и класс «Данные задачи реконфигурации». Класс «Алгоритмы»

содержит подклассы, относящиеся к общему описанию алгоритмов, например: GRASP, Имитация отжига, простой генетический алгоритм, случайный поиск (метод Монте-Карло). Поскольку онтология является дополняемой, на начальном этапе ее разработки ограничимся перечисленными алгоритмами.

Внутри каждого класса-алгоритма определены подклассы – признаки, в рамках которых могут быть классифицированы алгоритмы данного типа. Например, алгоритмы имитации отжига различают, в основном, по следующим признакам: закон изменения температуры, порождающее семейство распределений вероятностей, функция принятия нового состояния. Также подклассом класса «Имитация отжига» является «количество вычислений ЦФ», что является мерой вычислительной сложности применяемой реализации метода. На рис. 3 показан экземпляр алгоритма Имитации отжига с его отношениями с экземплярами подклассов, описывающих реализацию алгоритма имитации отжига.

Класс «Данные задачи реконфигурации» содержит подклассы:

- описание пользовательской задачи;
- описание фрагмента сети;
- число индивидуальных ограничений для устройств;
- число индивидуальных ЦФ;
- число общих ЦФ;
- число общих ограничений.

Подклассы класса «Описание пользовательской задачи» состоят из:

- амплитуды вычислительной сложности;
- амплитуды значений передаваемых объемов данных;
- диаметра графа;
- доминирующей вычислительной сложности;
- доминирующего объема передаваемых данных;
- количества подзадач в пакете;
- наличия информационных связей между задачами;
- наличия приоритетов для несвязанных задач;
- плотности связности графа.

Класс «Данные задачи реконфигурации» содержит экземпляры задач, которые, соответственно, связаны отношениями с экземплярами признаков задачи распределения нагрузки (рис. 4).

На рис. 5 показано графическое представление, где видно отношение между экземпляром постановки задачи и алгоритмом, который ее решает с высокой эффективностью.

Разработанная онтология является вспомогательным средством для последующего решения задачи выбора эффективного алгоритма в рамках поставленной общей задачи. Вполне очевидно, что при последующем дополнении разработанной онтологии по результатам проводимых экспериментальных исследований возможна реализация как логического вывода на основе продукционных правил, либо использования иных технологий ИИ в условиях неполной информации о задачах (например, классификации, а затем принятия решения по выбору эффективного алгоритма).

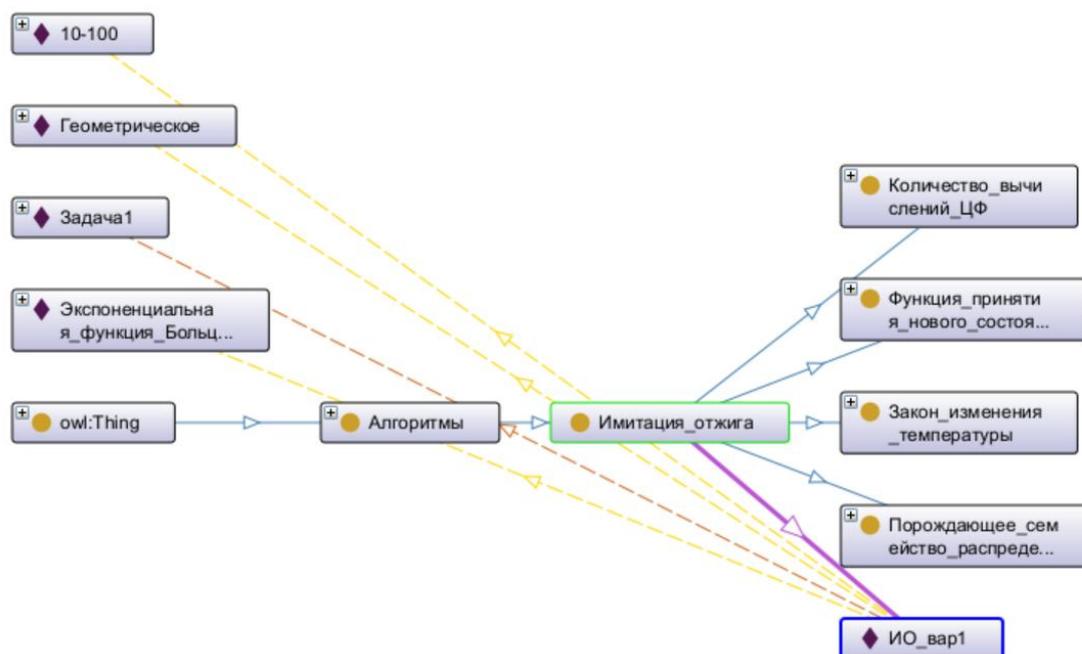


Рис. 3. Класс «Алгоритмы»

Fig. 3. "Algorithms" class

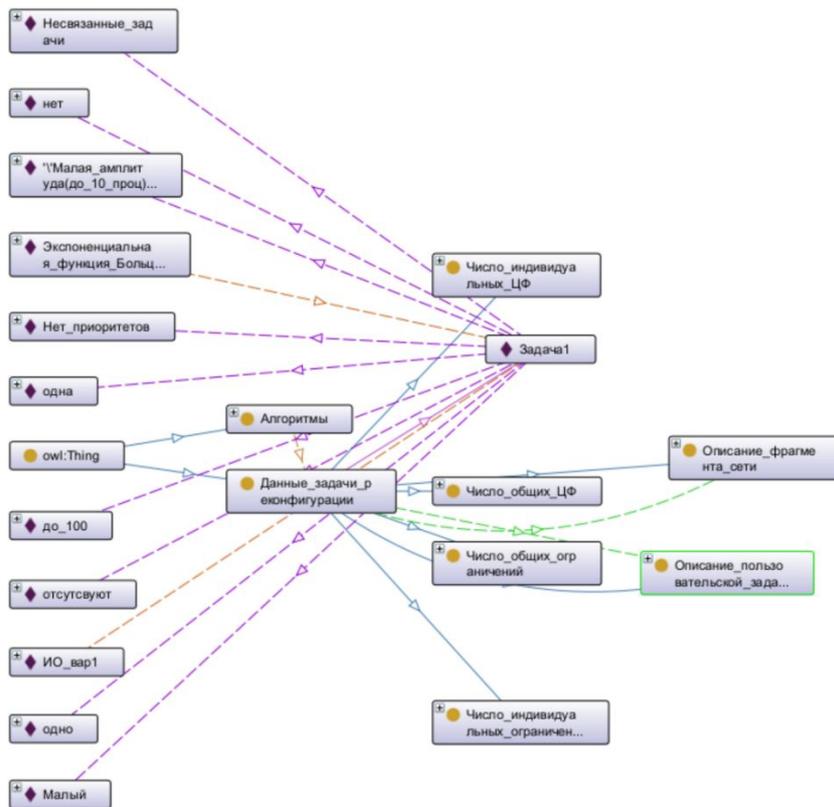


Рис. 4. Класс «Данные задачи реконфигурации»

Fig. 4. "Reconfiguration task data" class

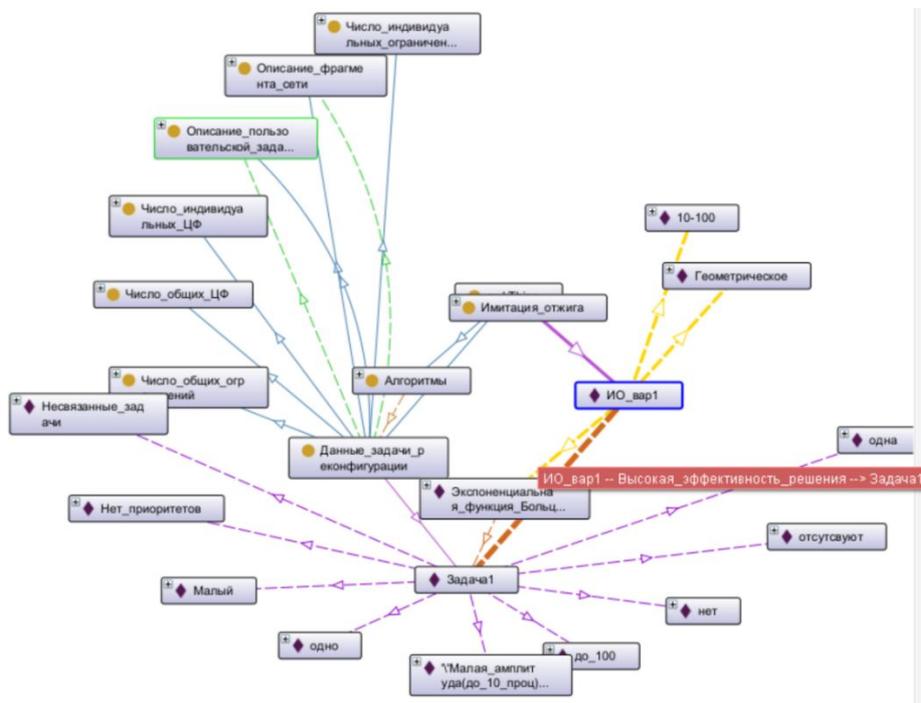


Рис. 5. Отношение «Высокая эффективность решения» между экземплярами класса «Имитация отжига» и «Описание пользовательской задачи»

Fig. 5. "High solution efficiency" relation between the "Simulated Annealing" individual and "User problem description" individual

Выводы

Данное исследование посвящено вопросам распределения вычислительных ресурсов при реализации распределенных вычислений в гетерогенных динамических распределенных вычислительных средах, частным случаем которых является среда туманных вычислений.

Ввиду отсутствия моделей задач распределения ресурсов, которые бы учитывали перечисленные аспекты систем распределенных вычислений и при этом не были бы частными реализациями, предложена модель, включающая как параметр процедуру изменения конфигурации системы.

Решение поставленной задачи возможно на основе использования «жадной» стратегии выбора наиболее эффективного алгоритма, а выбор, соответственно, возможен при наличии комплекта таковых.

Различие в эффективности используемых алгоритмов с различными параметрами относительно решения одной и той же задачи подтверждается экспериментальным исследованием – до 30% возможно улучшение качества решения, что, в целом, подтверждает актуальность формирования портфолио эффективных алгоритмов.

В рамках данного исследования была разработана онтологическая модель предметной области решения задач распределения ресурсов, в рамках которой, на основе отношений между экземплярами классов «Данные задачи реконфигурации» и «Алгоритмы» формируются комплекты алгоритмов, эффективных применительно к решению заданной задачи, что в дальнейшем позволит применять методы и технологии ИИ для быстрого выбора наиболее эффективного алгоритма.

Список литературы

1. Shams A., Shanjana S., Shaila A., Sabiha R., Mahfara H., Gandomi A. The Power of Internet of Things (IoT): Connecting the Dots with Cloud, Edge, and Fog Computing; 2023;70 p. <https://doi.org/10.48550/arXiv.2309.03420>
2. Hierarchical fuzzy-based Quality of Experience (QoE)-aware application placement in fog nodes // M. Sreenivasu, S. Satish, W. Rajeev, C. Raghavendra // *Software: Practice and Experience*. 2023;53(2):263-282. <https://doi.org/10.1002/spe.3147>
3. Dynamic Offloading in Flying Fog Computing: Optimizing IoT Network Performance with Mobile Drones / M. Wei, A. Khakimov, A. Ateya, M. Elaffendi, A. Muthanna, A. Abd El-Latif, M. Muthanna // *Drones*. 2023;7(10):622. <https://doi.org/10.3390/drones7100622>
4. Bhalekar P., Ingle S. A Comparative Study of Cluster, Grid and Cloud Computing // *International Journal of New Innovations in Engineering and Technology*, 2018;9(1):8-14. URL: <https://www.ijniet.org/wp-content/uploads/2018/11/2.pdf>.

5. Топорков В.В. Модели распределенных вычислений. М.: Физматлит, 2011. 320 с. URL: <https://biblioclub.ru/index.php?page=book&id=75957>.

6. Барский А.Б. Параллельные информационные технологии. М.: ИНТУИТ; БИНОМ. Лаборатория знаний, 2007. 503 с.

7. Методы и эвристики планирования в распределенных вычислениях с неотчуждаемыми ресурсами / В.В. Топорков, А.В. Бобченков, Д.М. Емельянов, А.С. Целищев // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика; 2014;3(2);43-62. URL: <https://www.mathnet.ru/rus/vyurv37>.

8. Метод повышения надежности за счет реконфигурации ресурсов в системах мониторинга и диагностики опасных природных явлений / Э.В. Мельник, М.В. Орда-Жигулина, Д.В. Орда-Жигулина, А.А. Родина // Известия ТулГУ. Серия: Технические науки; 2020;2:18-26. URL: <https://cyberleninka.ru/article/n/metod-povysheniya-nadezhnosti-za-schet-rekonfiguratsii-resursov-v-sistemah-monitoringa-i-dagnostiki-opasnyh-prirodnyh-yavleniy>.

9. Топорков В.В., Емельянов Д.М., Топоркова А.С. Анализ состояния исследований в области методов и алгоритмов планирования в грид и облачных вычислениях // ИТНОУ: информационные технологии в науке, образовании и управлении; 2018;2(6):26-31. URL: <https://cyberleninka.ru/article/n/analiz-sostoyaniya-issledovaniy-v-oblasti-metodov-i-algoritmov-planirovaniya-v-grid-i-oblachnyh-vychisleniyah>.

10. Sukhoroslov O. Scheduling of Workflows with Task Resource Requirements in Cluster Environments. In: Malyshkin V. (eds) Parallel Computing Technologies. PaCT 2023: Lecture Notes in Computer Science; 2023;14098:177-196. Springer, Cham. https://doi.org/10.1007/978-3-031-41673-6_14

11. Singh R.M., Awasthi L.K., Sikka G. Techniques for Task Scheduling in Cloud and Fog Environment: A Survey. In: Singh P., Sood S., Kumar Y., Paprzycki M., Pljonkin A., Hong WC. (eds) Futuristic Trends in Networks and Computing Technologies. FTNCT 2019. Communications in Computer and Information Science; 2020;1206:673-685. Springer, Singapore. https://doi.org/10.1007/978-981-15-4451-4_53

12. Nguyen B.M., Thi Thanh Binh H., Do Son B. Evolutionary algorithms to optimize task scheduling problem for the IoT based bag-of-tasks application in cloud-fog computing environment. Applied Science; 2019;9(9):1730. <https://doi.org/10.3390/app9091730>

13. Natesan G., Chokkalingam A. Task scheduling in heterogeneous cloud environment using mean grey wolf optimization algorithm. ICT Express; 2019;5(2):110-114. <https://doi.org/10.1016/j.ict.2018.07.002>

14. Narendrababu R.G., Phani K. S.: Modified ant colony optimization algorithm for task scheduling in cloud computing systems. In: Satapathy S.C., Bhateja V., Das S. (eds.) // Smart Intelligent Computing and Applications. SIST. Proceedings of the Second International

15. Bitam S., Zeadally S., Mellouk A. Fog computing job scheduling optimization based on bees swarm // *Enterprise Information Systems*; 2018;12(4);373–397. <https://doi.org/10.1080/17517575.2017.1304579>

16. Heuristic minconflicts optimizing technique for load balancing on fog computing. In: Xhafa F., Barolli L., Greguš M. (eds.) / M.B. Kamal, N. Javaid, S.A.A. Naqvi, H. Butt, T. Saif, M.D. Kamal // *The 10th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2018)*; 2019;23: 207–219. Springer, Cham. https://doi.org/10.1007/978-3-319-98557-2_19

17. Никишин К. И. Балансировка нагрузки данных в распределенной сети через прокси-сервер Nginx // *Известия Юго-Западного государственного университета*. 2022; 26(3): 98-111. <https://doi.org/10.21869/2223-1560-2022-26-3-98-111>.

18. Борзов Д.Б., Титов В.С., Басов Р.Г. Алгоритм и устройство планирования расписания загрузки процессоров в мультипроцессорных системах // *Известия Юго-Западного государственного университета*. 2019; 23(5): 161-174. <https://doi.org/10.21869/2223-1560-2019-23-5-161-174>

19. Борзов Д.Б., Масюков И.И. Планирование загрузки процессоров в мультипроцессорных системах критического назначения // *Известия Юго-Западного государственного университета*. 2018. Т. 22, № 6(81). С. 168-174.

20. Назаров С.В. Оптимизация вычислительного процесса бортовых вычислительных систем // *Известия Юго-Западного государственного университета*. 2018. Т. 22, № 2(77). С. 6-17.

References

1. Shams A, Shanjana S, Shaila A, Sabiha R, Mahfara H, Gandomi A. The Power of Internet of Things (IoT): Connecting the Dots with Cloud, Edge, and Fog Computing; 2023; 70 p. <https://doi.org/10.48550/arXiv.2309.03420>

2. Sreenivasu M, Satish S, Rajeev W, Raghavendra C. Hierarchical fuzzy-based Quality of Experience (QoE)-aware application placement in fog nodes. *Software: Practice and Experience*. 2023;53(2):263-282. <https://doi.org/10.1002/spe.3147>

3. Wei M, Khakimov A, Ateya A, Elaffendi M, Muthanna A, Abd El-Latif A, Muthanna M. Dynamic Offloading in Flying Fog Computing: Optimizing IoT Network Performance with Mobile Drones. *Drones*. 2023;7(10):622. <https://doi.org/10.3390/drones7100622>

4. Bhalekar P., Ingle S. A Comparative Study of Cluster, Grid and Cloud Computing. *International Journal of New Innovations in Engineering and Technology*, 2018;9(1):8-14. Available from: <https://www.ijniet.org/wp-content/uploads/2018/11/2.pdf>.

5. Toporkov V.V. *Modeli raspredelennykh vychislenii* [Distributed computing models]. Moscow, Phizmatlit Publ., 2011. 320 с. Available at: <https://biblioclub.ru/index.php?page=book&id=75957>.

6. Barskiy A.B. *Parallel'nye informatsionnye tekhnologii* [Parallel Information technologies]. Moscow, 2007. 503p.

7. Toporkov V.V., Bobchenkov A.V., Yemelianov D.M., Tselishev A.S. Metody i evristiki planirovaniya v raspredelennykh vychisleniyakh s neot-chuzhdaemyimi resursami [Methods and Heuristics of planning in distributed computing with the inalienable resources]. *Vestnik YuUrGU. Seriya: Vychislitel'naya matematika i informatika = Vestnik SUSU. Series: Computational Mathematics and Informatics*; 2014;3(2);43-62. Available at: <https://www.mathnet.ru/rus/vyurv37>.

8. Melnik E.V., Orda-Zhigulina M.V., Orda-Zhigulina D.V., Rodina A.A. Metod povysheniya nadezhnosti za schet rekonfiguratsii resursov v sistemakh monitoringa i diagnostiki opasnykh prirodnykh yavlenii [Reliability improvement method by means of resource reconfiguration in the systems of hazardous nature incidents monitoring and control]. *Izvestiya TulGU. Seriya: Tekhnicheskie nauki = Proceedings TSU. Series: Technical science*; 2020;2:18-26. Available at: <https://cyberleninka.ru/article/n/metod-povysheniya-nadezhnosti-za-schet-rekonfiguratsii-resursov-v-sistemah-monitoringa-i-diagnostiki-opasnyh-prirodnyh-yavleniy>.

9. Toporkov V.V., Yemelianov D.M., Toporkova A.C. Analiz sostoyaniya issledovaniya v oblasti metodov i algoritmov planirovaniya v grid i oblachnykh vychisleniyakh [The analysis of research state in the area of planning methods and algorithms in GRID and cloud computing]. *ITNOU: informatsionnye tekhnologii v nauke, obrazovanii i upravlenii = ITNOU: information technologies in science, education and management*; 2018;2(6):26-31. Available at: <https://cyberleninka.ru/article/n/analiz-sostoyaniya-issledovaniy-v-oblasti-metodov-i-algoritmov-planirovaniya-v-grid-i-oblachnykh-vychisleniyah>.

10. Sukhoroslov O. Scheduling of Workflows with Task Resource Requirements in Cluster Environments. In: Malyshev V. (eds) *Parallel Computing Technologies. PaCT 2023: Lecture Notes in Computer Science*; 2023;14098:177-196. Springer, Cham. https://doi.org/10.1007/978-3-031-41673-6_14

11. Singh R.M., Awasthi L.K., Sikka G. Techniques for Task Scheduling in Cloud and Fog Environment: A Survey. In: Singh P., Sood S., Kumar Y., Paprzycki M., Pljonkin A., Hong WC (eds) *Futuristic Trends in Networks and Computing Technologies. FTNCT 2019*.

Communications in Computer and Information Science; 2020;1206:673-685. Springer, Singapore. https://doi.org/10.1007/978-981-15-4451-4_53

12. Nguyen B.M., Thi Thanh Binh H., Do Son B. Evolutionary algorithms to optimize task scheduling problem for the IoT based bag-of-tasks application in cloud–fog computing environment. *Applied Science*; 2019;9(9):1730. <https://doi.org/10.3390/app9091730>

13. Natesan G., Chokkalingam A. Task scheduling in heterogeneous cloud environment using mean grey wolf optimization algorithm. *ICT Express*; 2019;5(2):110–114. <https://doi.org/10.1016/j.icte.2018.07.002>

14. Narendrababu R.G., Phani K.S. Modified ant colony optimization algorithm for task scheduling in cloud computing systems. In: Satapathy S.C., Bhateja V., Das S. (eds.) *Smart Intelligent Computing and Applications. SIST. Proceedings of the Second International Conference on SCI 2018*; 2019;104:357–365. Springer, Singapore. https://doi.org/10.1007/978-981-13-1921-1_36

15. Bitam S., Zeadally S., Mellouk A. Fog computing job scheduling optimization based on bees swarm. *Enterprise Information Systems*; 2018;12(4):373–397. <https://doi.org/10.1080/17517575.2017.1304579>

16. Kamal M.B., Javaid N., Naqvi S.A.A., Butt H., Saif T., Kamal M.D.: Heuristic min-conflicts optimizing technique for load balancing on fog computing. In: Xhafa F., Barolli L., Greguš M. (eds.). *The 10th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2018)*; 2019;23: 207–219. Springer, Cham. https://doi.org/10.1007/978-3-319-98557-2_19

17. Nikishin K.I. Load Balancer of Data in a Distributed Network via Nginx Proxy Server. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2022; 26(3): 98-111 (In Russ.). <https://doi.org/10.21869/2223-1560-2022-26-3-98-111>.

18. Borzov D. B., Titov V. S., Basov R.G. Algorithm and Planning Device for CPU Loading in Multiprocessor Systems. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2019, 23(5): 161-174 (In Russ.). <https://doi.org/10.21869/2223-1560-2019-23-5-161-174>

19. Borzov D.B., Masyukov I.I. Planning of Download Processors in Multiprocessor Systems of Critical Purpose. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*, 2018, vol. 22, no. 6(81), pp. 168-174.

20. Nazarov S.V. Optimization of Computing Process of the Onboard Computing Systems. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*, 2018, vol. 22, no. 2(77), pp. 6-17.

Информация об авторах / Information about the Authors

Клименко Анна Борисовна, кандидат технических наук, доцент кафедры Фундаментальной и прикладной математики, Институт информационных наук и технологий безопасности Российского государственного гуманитарного университета, г. Москва, Российская Федерация, e-mail: anna_klimenko@mail.ru, ORCID: <https://orcid.org/0000-0001-6527-8108>

Anna B. Klimenko, Cand. of Sci. (Engineering), Associate Professor, Fundamental and Applied Mathematics department, Institute of IT and Security Technologies of Russian State University for the Humanities, Moscow, Russian Federation, e-mail: anna_klimenko@mail.ru, ORCID: <https://orcid.org/0000-0001-6527-8108>

Алиева Эльвира Мурадovна, студент кафедры Фундаментальной и прикладной математики, Институт информационных наук и технологий безопасности Российского государственного гуманитарного университета, г. Москва, Российская Федерация, e-mail: anna_klimenko@mail.ru

Elvira M. Alieva, Student of Fundamental and Applied Mathematics department, Institute of IT and Security Technologies of Russian State University for the Humanities, Moscow, Russian Federation, e-mail: anna_klimenko@mail.ru

Сальников Андрей Евгеньевич, студент кафедры Фундаментальной и прикладной математики, Институт информационных наук и технологий безопасности Российского государственного гуманитарного университета, г. Москва, Российская Федерация, e-mail: anna_klimenko@mail.ru

Andrey Y. Salnikov, Student of Fundamental and Applied Mathematics department, Institute of IT and Security Technologies of Russian State University for the Humanities, Moscow, Russian Federation, e-mail: anna_klimenko@mail.ru