

УДК 004.021

Алгоритм планирования информационного обмена на основе бортовой ресурсосберегающей предобработки данных с использованием байесовского подхода

Басыров Александр Геннадьевич — докт. техн. наук, профессор, профессор кафедры информационно-вычислительных систем и сетей. E-mail: alexandrerb@mail.ru

Факхро Файруз — адъюнкт кафедры информационно-вычислительных систем и сетей. E-mail: fairouzarussia@gmail.com

Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Басыров А. Г., Факхро Ф. Алгоритм планирования информационного обмена на основе бортовой ресурсосберегающей предобработки данных с использованием байесовского подхода // Интеллектуальные технологии на транспорте. 2024. № 4 (40). С. 24–30. DOI: 10.20295/2413-2527-2024-440-24-30

Аннотация. Представлено обоснование, содержание и исследование алгоритма планирования ресурсосберегающего информационного обмена в условиях априорной неопределенности качества передаваемых данных. Приведена постановка задачи планирования информационного обмена с учетом анализа качества данных. Отражены результаты имитационного моделирования процессов информационного обмена с использованием рассматриваемого алгоритма. Предложенный алгоритм может быть использован для организации передачи данных в распределенных информационных системах реального времени при ограниченных энерго-временных ресурсах.

Ключевые слова: информационный обмен, априорная неопределенность качества данных

1.2.2 — математическое моделирование, численные методы и комплексы программ (технические науки), **2.3.1** — системный анализ, управление и обработка информации, статистика (технические науки)

Введение

Проблема оперативной передачи большого объема информации в распределенных информационных системах реального времени [1, 2] является одной из наиболее важных.

Путем ее решения является минимизация трафика на основе исключения передачи некачественных данных (например, некондиционных фотоснимков, дублированных информационных блоков, сбойных файлов данных и т. д.). Это можно выполнить предварительной обработкой (проверкой) информации перед ее передачей между узлами сети [3]. При этом сокращается объем передаваемых данных, но возни-

кает необходимость дополнительных временных затрат на подготовительную обработку.

Предварительная проверка качества информации обычно подразумевает проверку всех передаваемых данных. Наиболее жесткие ограничения на время информационного обмена приводят, наоборот, к передаче данных без проверки.

В статье предложен и исследован алгоритм, реализующий обоснованный компромисс между объемом проверяемых и передаваемых данных при ресурсных ограничениях (временных, энергетических и т. д.) на информационные процессы по обмену данными.

Постановка задачи планирования информационного обмена данными априорно-неопределенного качества

Постановка задачи информационного обмена данными априорно-неопределенного качества может быть сформулирована следующим образом.

Дано: множество $B = \{b_1, b_2, \dots, b_N\}$ блоков данных (БД), размещенных на первом объекте с априорно неизвестными значениями признаков $\pi(b_i)$ пригодности, где $\pi(b_i) \in \{0, 1\}$, $i = 1, \dots, N$, причем если $\pi(b_i) = 1$, то блок считается пригодным, в противном случае — дефектным.

Требуется: найти план U информационного процесса передачи подмножества $Y \in B$ блоков данных, находящихся на первом объекте, мощности $k \leq N$ такой, что:

$$U(B) = \text{Arg max}_{Y \in B} s(Y)$$

при ограничении $r(U) \leq R_{\text{доп}}$,

где $s(Y) = \sum_{j=1}^k \pi(b_j)$, $b_j \in Y$ — показатель качества информационного обмена, выражающийся количеством пригодных переданных блоков;

$r(U)$ — ресурс, затраченный на проверку и передачу блоков данных в соответствии с планом U ;

$R_{\text{доп}}$ — имеющийся ресурс на информационный обмен.

Значение r требуемого ресурса на реализацию информационного процесса (проверки и передачи данных) определяется выражением:

$$r = \eta r_t + r_s \left(\sum_1^{\eta} \pi(b_i) + \mu \right),$$

где η — количество проверенных блоков;

μ — количество переданных без проверки блоков;

$\sum_1^{\eta} \pi(b_i)$ — количество проверенных блоков, оказавшихся по результатам проверки исправными (рис. 1).

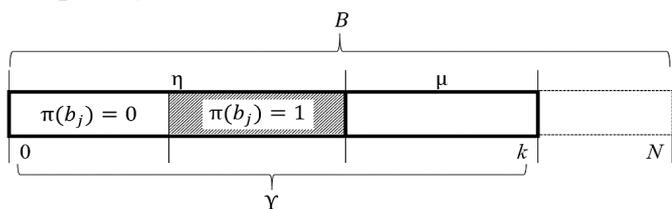


Рис. 1. Схема распределения блоков данных

План U информационного обмена, включающий процессы проверки и передачи блоков данных, в описанных выше условиях может быть формализован вектором $U = u_{\beta_i, k}$, $k \leq N$, задающим последовательность проверки и/или передачи блоков данных, где u_{β_i} — действие с блоком данных b_{β_i} в виде:

$$u_{\beta_i} = \begin{cases} 1 & \text{— проверить } b_{\beta_i}\text{-й блок данных, и если } \pi(b_{\beta_i}) = 1, \text{ то передать его, иначе — не передавать;} \\ 0 & \text{— передать } b_{\beta_i}\text{-й блок данных без проверки.} \end{cases}$$

Количество k элементов вектора U определяется количеством η проверенных и μ переданных блоков, а также числом пригодных из проверенных блоков.

Стратегия распределенного ресурсосберегающего информационного обмена при неопределенности качества информации [4] заключается в передаче проверенных блоков данных, оказавшихся по результатам проверок пригодными, и оставшихся — без проверки ($\eta \geq 0, \mu \geq 0$) с учетом достаточности ресурса на проверку и передачу блоков. Очевидно, что такая стратегия анализа и передачи блоков данных имеет смысл только в случае, когда затраты на проверку одного блока данных меньше затрат на его передачу ($r_t < r_s$). В противном случае предварительная проверка блоков нецелесообразна.

Алгоритм планирования информационного обмена с учетом анализа качества данных

Для решения поставленной задачи применение методов статистического контроля качества продукции [5] невозможно, так как необходимо учитывать ресурсоемкость операций контроля качества данных. Кроме того, имеет место ограниченный и, возможно, весьма малый объем передаваемых данных.

Рассмотрим два подхода к организации информационного обмена с предварительным анализом качества данных.

Первый подход (статический) предполагает передачу БД либо с предварительной проверкой всех БД ($u_i = 1, \forall i = 1, 2, \dots, k$), либо без

предварительной их проверки ($u_i = 0, \forall i = 1, 2, \dots, k$). Учитывая полную неопределенность качества данных, вероятность пригодности любого БД априори принимается за 0,5. При этом в первом случае количество элементов вектора U равно:

$$k = \frac{R}{\eta r_t + r_s \sum_1^\eta \pi(b_i)},$$

а во втором случае:

$$k = \left\lfloor \frac{R}{r_s} \right\rfloor$$

где η — количество проверенных БД;

$\lfloor x \rfloor$ — «ближайшее целое, меньшее или равное x ».

При этом предполагается, что распределение пригодных БД соответствует закону Бернулли [6] с вероятностью пригодности каждого БД, равной p .

При статическом решении не проверять БД ($u_i = 0$) математическое ожидание количества пригодных БД составит:

$$m_1 = \frac{R}{r_s} p,$$

а при решении с проверкой всех БД ($u_i = 1$) математическое ожидание количества пригодных БД составит:

$$m_2 = \frac{R}{r_t + p r_s} p.$$

Тогда целесообразность решения «проверить все БД» обоснована при условии $m_2 > m_1$, то есть при:

$$\frac{R}{r_t + p r_s} p > \frac{R}{r_s} p,$$

что равносильно условию:

$$p < 1 - \frac{r_t}{r_s},$$

а с учетом того, что $p = 1/2$ — условию:

$$\frac{r_t}{r_s} < \frac{1}{2}.$$

При втором подходе (динамическом) предполагается передача части БД с предварительной проверкой, а оставшейся части — без проверки.

Здесь также полагается, что на любом i -м шаге информационного обмена распределение пригодных БД соответствует закону Бернулли с вероятностью

пригодности каждого БД, равной p_i , однако, в отличие от предыдущего подхода, значение p_i не является константой, а изменяется по результатам проверки очередного БД.

Оценим математическое ожидание количества пригодных БД на основе использования априорной и апостериорной вероятности по Байесу [7].

На очередном i -м шаге требуется принять решение u_i относительно того, проверять очередной БД ($u_i = 1$) или нет ($u_i = 0$).

Если принято решение не проверять оставшиеся БД ($u_i = 0$), то математическое ожидание количества пригодных БД составит:

$$m_1 = \frac{R}{r_s} p_i,$$

где p_i — априорная вероятность пригодности i -го БД.

Если же принято решение проверить очередной БД ($u_i = 1$), то математическое ожидание количества пригодных непроверенных БД составит:

$$m_2 = p_i \left(1 + \frac{R - r_t - r_s}{r_s} p_{1i} \right) + (1 - p_i) \left(\frac{R - r_t}{r_s} p_{0i} \right).$$

Здесь p_{1i} — апостериорная вероятность пригодности одного БД в случае, если очередной i -й проверенный БД окажется пригодным, а p_{0i} — апостериорная вероятность пригодности одного БД в случае, если он окажется непригодным.

Эти апостериорные вероятности можно найти из следующих соотношений:

$$p_{0i} = \frac{p_i(i+1)}{i+2}, \quad p_{1i} = \frac{1 + p_i(i+1)}{i+2}.$$

Тогда решение $u_i = 1$ обоснованно при $m_2 > m_1$, то есть при:

$$p_i \left(r_s - (r_t + r_s) \frac{1 + p_i(i+1)}{i+2} \right) - (1 - p_i) r_t \frac{p_i(i+1)}{i+2} > 0.$$

Упрощая это неравенство, получим условие, при котором целесообразна проверка очередного БД в виде:

$$p_i < 1 - \frac{r_t(i+2)}{r_s(i+1)}.$$

Ниже представлен комплексный алгоритм, состоящий из двух алгоритмов информационного обмена с учетом анализа качества данных, реализующих статический (алгоритм 1) и динамический (алгоритм 2) подходы к организации информационного обмена.

Алгоритм 1. Статический анализ пригодности данных:

Шаг 1. Начало.

Шаг 2. $i := 1, s := 0, r := 0$.

Шаг 3. Если $r < R \wedge i \leq N$, то переход на шаг 4, иначе — на шаг 10.

Шаг 4. Если $\frac{r_i}{r_s} < \frac{1}{2}$, то переход на шаг 5, иначе — на шаг 6.

Шаг 5. $r := r + r_i + r_s \cdot (b_i)$. Если $\pi(b_i) = 1$, то переход на шаг 7, иначе — на шаг 8.

Шаг 6. $r := r + r_s$.

Шаг 7. Передать i -й БД.

Шаг 8. $s := s + \pi(b_i)$.

Шаг 9. $i := i + 1$. Переход на шаг 3.

Шаг 10. Конец.

Алгоритм 2. Динамический анализ пригодности данных:

Шаг 1. Начало.

Шаг 2. $i := 1, s := 0, r := 0$.

Шаг 3. Если $r < R \wedge i \leq N$, то переход на шаг 4, иначе — на шаг 12.

Шаг 4. Если $p < 1 - r_i(i+2) / r_i(i+1)$, то переход на шаг 5, иначе — на шаг 8.

Шаг 5. $r := r + r_i + r_s \cdot (b_i)$.

Шаг 6. $p := (\pi(b_i) + p(i+1)) / (i+2)$.

Шаг 7. Если $\pi(b_i) = 1$, то переход на шаг 9, иначе — на шаг 10.

Шаг 8. $r := r + r_s$.

Шаг 9. Передать i -й БД.

Шаг 10. $s := s + \pi(b_i)$.

Шаг 11. $i := i + 1$. Переход на шаг 3.

Шаг 12. Конец.

Вычислительная сложность обоих алгоритмов составляет $O(N)$. В результате работы алгоритмов значение переменной s будет равно количеству переданных пригодных БД.

Следует заметить, что на основе вышеизложенного представленные алгоритмы позволяют спланировать информационный обмен с учетом анализа качества данных и ресурсных ограничений.

Результаты моделирования планирования информационного обмена с учетом анализа качества данных

Для оценивания результативности предлагаемого алгоритма было проведено имитационное моделирование процессов планирования информационным обменом.

Исходными данными для моделирования являлись:

N — количество передаваемых БД;

r_s — ресурсные затраты на передачу одного БД;

r_i — ресурсные затраты на проверку одного БД;

R — имеющийся для информационного обмена ресурс.

Результативность применения предлагаемого алгоритма выражена средним значением γ отношения количества переданных пригодных БД к максимально возможному при заданных ресурсных ограничениях количеству пригодных БД.

В ходе моделирования на основе генератора случайных чисел задавались значения вектора $\pi(b_i) = 1, \dots, N$ пригодности БД.

На рис. 2 представлены зависимости показателя результативности γ алгоритмов с проверкой данных, без проверки данных и алгоритма статического анализа пригодности данных (алгоритм 1) от соотношения затрат на проверку и передачу данных для 20 БД при доступном ресурсе, достаточном для передачи всех пригодных БД.

На каждом шаге имитационного моделирования значение вектора пригодности БД менялось и приняло все возможные значения, а общее число шагов моделирования составило 2^N .

На рис. 3 представлены зависимости показателя результативности γ алгоритмов с проверкой данных, без проверки данных и алгоритма динамического анализа пригодности данных (алгоритм 2) от соотношения затрат на проверку и передачу данных для 20 БД при доступном ресурсе, достаточном для передачи всех пригодных БД.

При моделировании предполагалось, что все дефектные БД размещены «кучно» в одной области, причем их количество принимало все возможные значения.

Отметим, что в случае динамического анализа пригодности данных доля непригодных БД априори неизвестна.

При этом алгоритм 2 обеспечил значения показателя качества, близкие к лучшим из двух других алгоритмов.

Таким образом, из анализа проведенных исследований следует, что предложенный алгоритм может быть использован на практике при планировании информационного обмена в условиях ре-

сурсных ограничений и неопределенности качества информации.

Заключение

Приведенный алгоритм имеет практическую направленность на обоснование аппаратно-программных средств обеспечения передачи данных в мобильных устройствах при ограниченном энерго-временном ресурсе.

В отличие от известных [8–10], он реализует байесовский подход к оцениванию вероятности передачи пригодных БД, что позволяет сократить объем передаваемых данных без существенной потери их значимости.

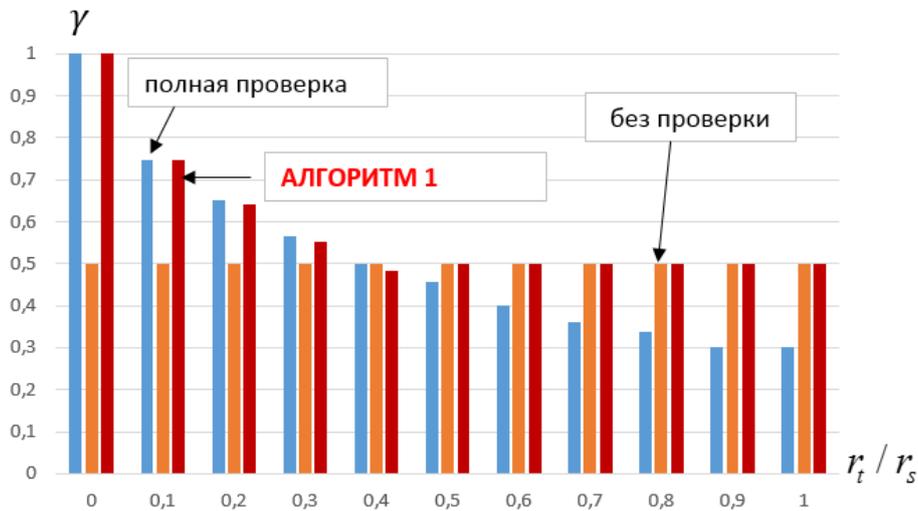


Рис. 2. Результативность статического анализа данных при изменении соотношения затрат на проверку и передачу БД

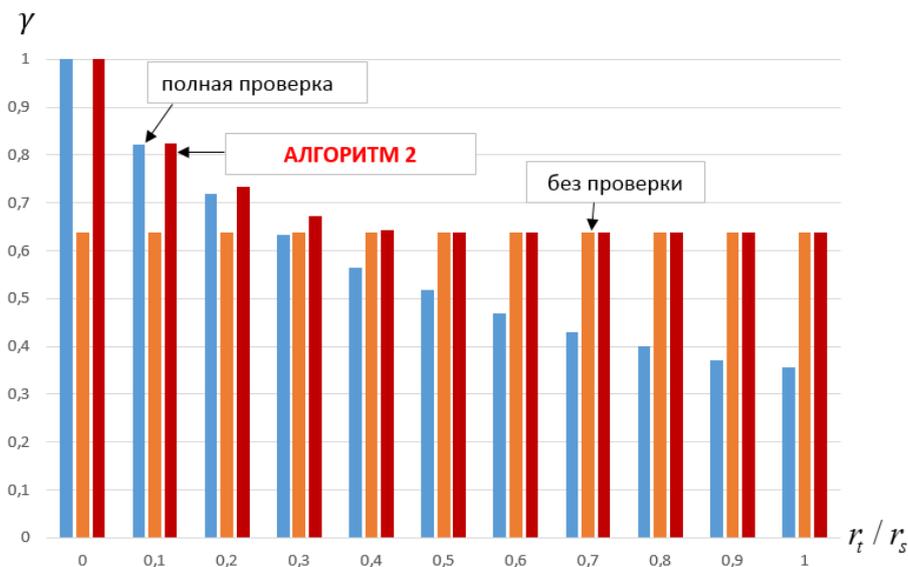


Рис. 3. Результативность динамического анализа данных при изменении доли непригодных БД

Результаты проведенного имитационного моделирования процессов информационного обмена с использованием предложенных алгоритмов свидетельствуют о том, что статический анализ пригодности данных целесообразно применять при

полной априорной неопределенности качества передаваемой информации. Динамический анализ актуален в ситуации, когда непригодные (сбойные) БД расположены подряд в общей последовательности блоков данных, что часто имеет место на практике.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Луканов А. С. Системы реального времени: учебное пособие // Самара: Издательство Самарского университета, 2020. 156 с.
2. Distributed Systems: Concepts and Design / G. Coulouris [et al.]. Fifth edition. Addison-Wesley, 2011.
3. Балухто А. Н., Ключников В. Ю., Хартов В. В. Интеллектуальная обработка целевой информации на борту космических аппаратов многоспутниковых систем // Космонавтика и ракетостроение. 2020. № 3 (114). С. 49–63.
4. Басыров А. Г., Лупашко М. Н. Алгоритмы ресурсосберегающих информационных процессов на основе предварительного анализа качества данных // Технологии, алгоритмы и программы для решения прикладных задач кибербезопасности, помехозащищенности и информационного обеспечения. Вып. 5 (39) / под общ. ред. В. А. Овчарова. СПб.: ВКА имени А. Ф. Можайского, 2022. С. 157–163.
5. ГОСТ Р 50779.12–2021. Статистические методы. Статистический контроль качества. Методы случайного отбора выборок штучной продукции. М.: Стандартинформ, 2021. 18 с.
6. Вентцель Е. С., Овчаров Л. А. Теория вероятностей и ее инженерные приложения: учебное пособие. 5-е изд. М.: КНОРУС, 2013. 480 с.
7. Тимофеев В. В., Лупашко М. Н., Степанов И. В. Алгоритм планирования энергосберегающих информационных процессов с использованием байесовского подхода // Технологии, алгоритмы и программы для решения прикладных задач кибербезопасности, помехозащищенности и информационного обеспечения. Вып. 4(38) / под общ. ред. В. А. Овчарова. СПб.: ВКА имени А. Ф. Можайского, 2021. С. 168–174.
8. Басыров А. Г., Кузнецов В. В., Лупашко М. Н. Алгоритмы ресурсосберегающего автономного контроля массива данных наблюдения в беспилотных летательных аппаратах // Вестник Российского нового университета. Серия «Сложные системы: модели, анализ и управление». 2022. Вып. 3. С. 3–10.
9. Басыров А. Г., Калюжный А. В., Хомоненко А. Д. Алгоритм двуххранцевой упаковки с ограниченным межграничным обменом // Интеллектуальные технологии на транспорте. 2022. № 1 (29). С. 29–33.
10. Калюжный А. В., Зыкова С. С., Терехов В. Г. Алгоритм поиска кратчайшего пути между подвижными объектами транспортной сети // Интеллектуальные технологии на транспорте. 2020. № 3. С. 5–10.

Дата поступления: 02.11.2024

Решение о публикации: 02.11.2024

An Algorithm for Planning Information Exchange Based on On-Board Resource-Saving Data Preprocessing Using the Bayesian Approach

Alexander G. Basyrov — Dr. Sci. in Engineering, Professor of the Department of Information and Computing Systems and Networks. E-mail: alexanderbas@mail.ru

Fakhro Fairouz — Advanced Student of the Department of Information and Computing Systems and Networks. E-mail: fairouzarussia@gmail.com

Mozhaisky Military Aerospace Academy, 13, Zhdanovskaya str., St. Petersburg, 197198, Russia

For citation: Basyrov A. G., Fakhro F. An Algorithm for Planning Information Exchange Based on On-Board Resource-Saving Data Preprocessing Using the Bayesian Approach // Intellectual Technologies on Transport. 2024. № 4 (40). Pp. 24–30. DOI: 10.20295/2413-2527-2024-440-24-30. (In Russian)

Abstract. *The substantiation, content and research of the algorithm for planning resource-saving information exchange in conditions of a priori uncertainty of the quality of transmitted data is presented. The formulation of the task of planning information exchange is given, taking into account the analysis of data quality. The results of simulation modeling of information exchange processes using the considered algorithm are reflected. The proposed algorithm can be used to organize data transmission in distributed real-time information systems with limited energy and time resources.*

Keywords: *information exchange, a priori uncertainty of data quality*

REFERENCES

1. Lukanov A. S. Sistemy real'nogo vremeni: uchebnoe posobie. Samara: Izdatel'stvo Samarskogo universiteta, 2020. 156 s. (In Russian)
2. Distributed Systems: Concepts and Design / G. Coulouris [et al.]. Fifth edition. Addison-Wesley, 2011.
3. Baluhto A. N., Klyushnikov V. Yu., Hartov V. V. Intel'ktual'naya obrabotka celevoj informacii na borte kosmicheskikh apparatov mnogosputnikovykh sistem // Kosmonavtika i raketostroenie. 2020. No. 3 (114). S. 49–63. (In Russian)
4. Basyrov A. G., Lupashko M. N. Algoritmy resursoberegayushchih informacionnykh processov na osnove predvaritel'nogo analiza kachestva dannykh // Tekhnologii, algoritmy i programmy dlya resheniya prikladnykh zadach kiberneticheskoy bezopasnosti, pomekhozashchishennosti i informacionnogo obespecheniya. Vyp. 5 (39) / pod obshch. red. V. A. Ovcharova. SPb.: VKA imeni A. F. Mozhajskogo, 2022. S. 157–163. (In Russian)
5. GOST R 50779.12–2021. Statisticheskie metody. Statisticheskij kontrol' kachestva. Metody sluchajnogo otbora vyborok shtuchnoj produkcii. M.: Standartinform, 2021. 18 s. (In Russian)
6. Ventcel' E. S., Ovcharov L. A. Teoriya veroyatnostej i ee inzhenernye prilozheniya: uchebnoe posobie. 5-e izd. M.: KNORUS, 2013. 480 s. (In Russian)
7. Timofeev V. V., Lupashko M. N., Stepanov I. V. Algoritm planirovaniya energosberegayushchih informacionnykh processov s ispol'zovaniem bajesovskogo podhoda // Tekhnologii, algoritmy i programmy dlya resheniya prikladnykh zadach kiberneticheskoy bezopasnosti, pomekhozashchishennosti i informacionnogo obespecheniya. Vyp. 4(38) / pod obshch. red. V. A. Ovcharova. SPb.: VKA imeni A. F. Mozhajskogo, 2021. S. 168–174. (In Russian)
8. Basyrov A. G., Kuznecov V. V., Lupashko M. N. Algoritmy resursoberegayushchego avtonomnogo kontrolya massiva dannykh nablyudeniya v bespilotnykh letatel'nykh apparatah // Vestnik Rossijskogo novogo universiteta. Seriya "Slozhnye sistemy: modeli, analiz i upravlenie". 2022. Vyp. 3. S. 3–10. (In Russian)
9. Basyrov A. G., Kalyuzhnyj A. V., Homonenko A. D. Algoritm dvuhrancevoj upakovki s ogranichenym mezhhrancevym obmenom // Intel'ktual'nye tekhnologii na transporte. 2022. No. 1 (29). S. 29–33. (In Russian)
10. Kalyuzhnyj A. V., Zykova S. S., Terekhov V. G. Algoritm poiska krachajshogo puti mezhdru podvizhnymi ob'ektami transportnoj seti // Intel'ktual'nye tekhnologii na transporte. 2020. No. 3. S. 5–10. (In Russian)

Received: 02.11.2024

Accepted: 02.11.2024