

УДК 004.94

СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ПОЛИМОДАЛЬНЫМИ ПОТОКАМИ

С.А. Жмылёв^а, Т.И. Алиев^а

^а Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: korg@cs.ifmo.ru

Информация о статье

Поступила в редакцию 19.02.18, принята к печати 26.03.18

doi: 10.17586/2226-1494-2018-18-3-473-478

Язык статьи – русский

Ссылка для цитирования: Жмылёв С.А., Алиев Т.И. Системы массового обслуживания с полимодальными потоками // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 3. С. 473–478. doi: 10.17586/2226-1494-2018-18-3-473-478

Аннотация

При моделировании систем массового обслуживания традиционно предполагается, что распределение интервалов между поступающими в систему заявками имеет унимодальный вид. В то же время на практике встречаются системы с полимодальным распределением интервалов, имеющим две и более моды. При исследовании таких систем полимодальное распределение обычно заменяется унимодальным с совпадающими числовыми моментами, что упрощает процесс моделирования, но вносит методическую погрешность в получаемые результаты. Предметом исследования является анализ свойств полимодальных потоков и оценка величины этой погрешности. В качестве метода исследования предложено имитационное моделирование в среде AnyLogic, позволившее выявить зависимость погрешности от загрузки системы. Выявлено, что в высоконагруженных системах вносимая заменой погрешность не превышает 15%, но с уменьшением загрузки системы погрешность увеличивается и может достигать сотен процентов. В ходе многочисленных экспериментов выявлено, что значение коэффициента вариации интервалов между поступающими заявками в полимодальном потоке с увеличением межмодового интервала стремится к значению, не превышающему единицу. Практическая значимость работы состоит в упрощении процесса проектирования высокозагруженных вычислительных систем с использованием имитационных и аналитических моделей за счет замены полимодальных потоков унимодальными. При этом обеспечивается заданная точность расчета характеристик.

Ключевые слова

полимодальное распределение, система с очередями, аппроксимация полимодальных распределений, числовой момент, мультиэкспоненциальное распределение

QUEUE SYSTEMS WITH POLYMODAL QUERY FLOWS

S.A. Zhmylev^а, T.I. Aliev^а

^а ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: korg@cs.ifmo.ru

Article info

Received 19.02.18, accepted 26.03.18

doi: 10.17586/2226-1494-2018-18-3-473-478

Article in Russian

For citation: Zhmylev S.A., Aliev T.I. Queue systems with polymodal query flows. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 3, pp. 473–478 (in Russian). doi: 10.17586/2226-1494-2018-18-3-473-478

Abstract

In queuing systems modeling it is traditionally assumed that the distribution of the inter-arrival time between requests entering the system is unimodal. At the same time, in practice, systems with a multimodal inter-arrival time distribution having two or more modes are encountered. The multimodal distribution is usually replaced by a unimodal one with appropriate numerical moments, which simplifies the modeling process, but introduces a methodical error in the obtained results. The subject of the study is the analysis of properties of multimodal flows and the estimation of such an error. Simulation modeling in the AnyLogic environment is used as a re-search method provided to discover the error dependence on the system load. It was revealed that in the high-load systems the error introduced by the replacement is not higher than 15%, but with a decrease in the system load, the error increases and can reach hundreds of percent. In the course of extensive experiments, it was found out that with an increase in the inter-arrival time the value of the variation coefficient between incoming queries in a polymodal flow tends to a constant value not greater than unity. The practical importance of the work

lies in simplification of design process for high-load computing systems with the use of simulation and analytical models by replacing the polymodal flows with unimodal providing required accuracy of calculations.

Keywords

polymodal distribution, queue system, polymodal distribution approximation, statistical moment, multi-exponential distribution

Введение

При исследовании систем с очередями для построения модели потока обычно измеряются интервалы между происходящими событиями, затем вычисляются математическое ожидание и коэффициент вариации, и на основе этих характеристик выполняется аппроксимация потока заявок унимодальным распределением [1–3]. В то же время, если построить гистограмму реального распределения, то она может иметь полимодальный вид. Примерами реальных систем, для которых характерна такая гистограмма распределения, могут служить беспроводная вычислительная сеть с централизованным контролем доступа к среде передачи или вычислительная сеть с подтверждением приема сетевых пакетов [4–7]. И в том, и в другом случаях передаваемое сообщение разделяется на сетевые пакеты [8, 9]. Отправка каждого сетевого пакета требует также передачи служебного сетевого пакета меньшего размера [10–12]. В связи с этим в передающее оборудование пакеты поступают с распределением, имеющим две моды. Возникает вопрос правомочности аппроксимации реального распределения с помощью унимодальной функции. Материалов, посвященных данному вопросу, при обзоре предметной области не было найдено. Цель работы состоит в выявлении свойств систем с полимодальным распределением интервалов поступления заявок и оценке погрешности, возникающей при замене полимодального распределения унимодальным. Параметры используемого в модели распределения получены в результате анализа потоков данных в вычислительной сети с подтверждением приема сетевых пакетов.

Постановка задачи

Рассмотрим телекоммуникационную систему, работающую по протоколу передачи данных с подтверждением, состоящую из трех последовательно соединенных вычислительных узлов. Узлы функционируют под управлением операционной системы GNU/Linux версии Debian 8.4. Между узлами происходит обмен сообщениями размером 8,5 МБ. На основе результатов, полученных в процессе измерений характеристик потоков пакетов, построена гистограмма распределения времени (в миллисекундах) между поступлениями сетевых пакетов в промежуточный узел при установленных в ядре операционной системы настройках `net.ipv4.tcp_wmem` и `net.ipv4.tcp_rmem` в значение «1500, 1500, 1500» (рис. 1).

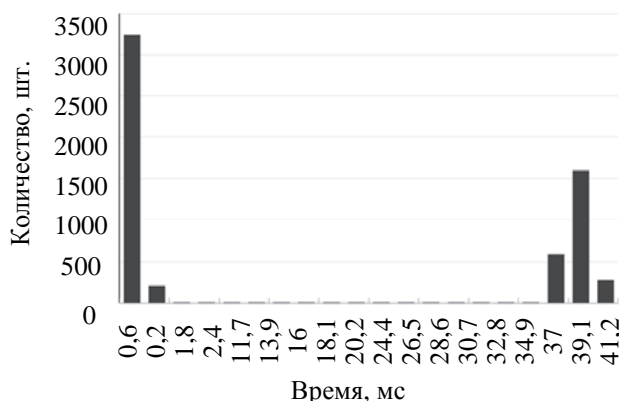


Рис. 1. Гистограмма распределения интервалов между сетевыми пакетами

Очевидно, функция плотности распределения имеет две моды. Такой поток заявок можно представить в модели как унимодальной функцией, так и полимодальной, имеющей две моды. Пусть имеется система массового обслуживания с экспоненциальным распределением времени обслуживания заявок. Входной поток задается полимодальной функцией с различным числом компонент. Рассмотрим варианты замены полимодального распределения унимодальным и погрешность, вносимую в результаты моделирования вследствие такой замены.

Анализ полимодальных потоков

Полимодальную функцию можно получить путем композиции унимодальной функции $f(x)$ и ее смещенного варианта $f(x - D)$. В работе рассматриваются распределения, функции которых имеют две и три моды и являются мультиэкспонентами (ввиду удобства построения используются гипо- и гиперэкспоненты). Каждая мода полимодальной функции смещается относительно предыдущей на равные интервалы D по оси абсцисс, т.е. $D_i = D_{i-1} + D = iD$; ($i = 1, 2, \dots$); $D_0 = 0$. Например, распределение с интен-

сивностью λ , состоящее из композиции двух экспоненциальных функций, каждая из которых определяет значение случайной величины с вероятностью p , имеет плотность $f(x)$ и математическое ожидание M :

$$f(x) = \begin{cases} 0, & x < 0 \\ p\lambda e^{-\lambda x}, & 0 \leq x < D \\ p\lambda e^{-\lambda x} + (1-p)\lambda e^{-\lambda(x-D)}, & x \geq D \end{cases}; \quad M = D(1-p) + \frac{1}{\lambda}.$$

Характер полимодального распределения, полученного по этим формулам с разными значениями параметров p, D и λ , а именно: $p_1 = p_2 > p_3$; $D_3 > D_2 > D_1$; $\lambda_3 = \lambda_1 > \lambda_2$, имеет вид кривых 1, 2 и 3 (рис. 2).

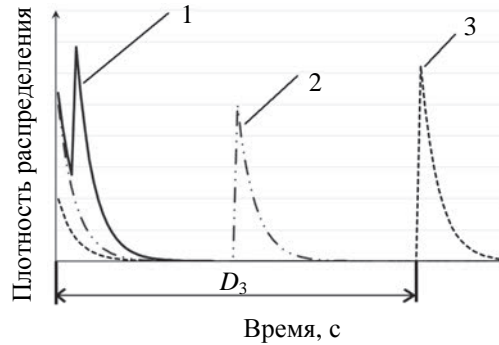


Рис. 2. Примеры двухмодовых распределений

Для выявления свойств рассматриваемых полимодальных распределений построены имитационные модели потоков, составленные из функций, имеющих равные математические ожидания. Имитационные эксперименты выполнялись в среде AnyLogic 7.0.2 Professional в предположениях о детерминированном и экспоненциальном распределении времени обслуживания. Модель представляет собой один или несколько приборов, обслуживающих поступающие заявки, и накопитель, в который помещается заявка, поступившая, когда все приборы заняты обслуживанием. Полимодальная функция задает распределение времени между поступлениями заявок в модель. При этом варьировались следующие параметры: величина смещения каждой моды D , создаваемая потоком нагрузка u , распределение времени обслуживания поступающих заявок $f(\tau)$. В качестве исследуемых характеристик рассматривались среднее время пребывания заявок U в модели, загрузка прибора ρ , средняя длина очереди l и усредненная максимальная длина очереди l_{\max} , позволяющая оценить верхнюю границу емкости накопителя проектируемой системы. Для исключения влияния переходных процессов на измеряемые величины через системы пропускалось 500 тыс. заявок, что обеспечило отклонение характеристик систем не более 0,1%.

Аппроксимация полимодальных распределений

Аппроксимация вероятностных распределений по двум числовым моментам – математическому ожиданию M и второму начальному моменту – требует выбора аппроксимирующей функции на основе коэффициента вариации v . Для распределений с $v > 1$ можно использовать гиперэкспоненциальное распределение, составленное из двух экспоненциальных распределений с математическими ожиданиями M_1 и M_2 соответственно, одно из которых формирует значение случайной величины с вероятностью q , а второе – с вероятностью $(1 - q)$. В этом случае формулы для нахождения параметров гиперэкспоненциального распределения [13, 14] имеют следующий вид:

$$q \leq \frac{2}{v^2 + 1}; \quad M_1 = M \left[1 + \sqrt{\frac{1-q}{2q}}(v^2 - 1) \right]; \quad M_2 = M \left[1 - \sqrt{\frac{q}{2(1-q)}}(v^2 - 1) \right].$$

Для аппроксимации полимодальных распределений с коэффициентом вариации $0 < v < 1$ использовались гипоекспоненциальные распределения, представляемые суммой k_1 экспоненциальных распределений с математическими ожиданиями M_1 и k_2 экспоненциальных распределений с математическими ожиданиями M_2 . При этом $k = k_1 + k_2$. Формулы для нахождения параметров в этом случае имеют следующий вид [13, 14]:

$$k \geq \frac{1}{v^2}; \quad M_1 = \frac{M}{k} \left[1 + \sqrt{\frac{k_2}{k_1}}(kv^2 - 1) \right]; \quad M_2 = \frac{M}{k} \left[1 - \sqrt{\frac{k_1}{k_2}}(kv^2 - 1) \right].$$

Результаты экспериментов

В ходе имитационного моделирования было получено множество гистограмм полимодальных распределений. Каждое распределение было аппроксимировано мультиэкспоненциальной функцией по двум

числовым моментам. Характер исходной функции плотности распределения, представляемой гистограммой, существенно отличается от характера аппроксимирующей функции, представляемой кривой (рис. 3).

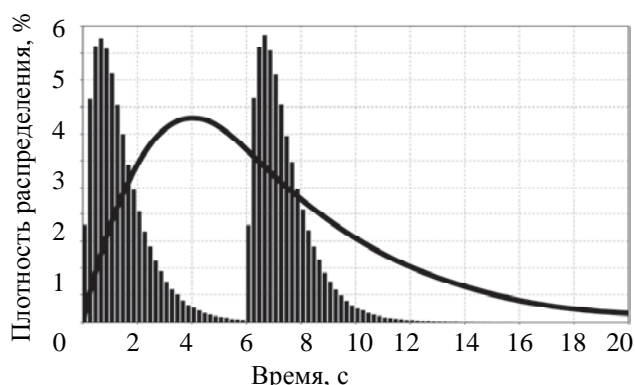


Рис. 3. Аппроксимация полимодального распределения

Загрузка	0,3		0,7		0,95	
	l , шт	$0,964 \pm 0,007$	124%	$2,655 \pm 0,08$	14%	$22,20 \pm 4,90$
	$0,429 \pm 0,005$	$2,33 \pm 0,044$		$20,22 \pm 3,85$		
l_{\max} , шт	$17,8 \pm 1,77$	69%	$36,8 \pm 3,45$	10%	$179,1 \pm 29,76$	25%
	$10,5 \pm 0,88$		$33,5 \pm 1,49$		$142,8 \pm 20,65$	
U , с	$14,684 \pm 0,26$	37%	$63,81 \pm 2,07$	9%	$553,75 \pm 122,2$	9%
	$10,71 \pm 0,14$		$58,1585 \pm 1,12$		$504,5 \pm 96,15$	

Таблица. Сравнение характеристик систем с полимодальными и унимодальными потоками

В таблице показаны результаты сравнения систем с полимодальными и аппроксимирующими его унимодальными мультиэкспоненциальными потоками. Уровень доверия результатов составляет 95%. Интервалы между поступлением заявок в потоке имеют математическое ожидание $24,933 \pm 0,02$, коэффициент вариации $1,00015 \pm 0,0004$. Полимодальный поток составлен из двух гиперэкспоненциальных компонент с коэффициентом вариации, равным 5, расположенных на расстоянии 45,9.

Полученные значения характеристик для полимодального и унимодального потоков представлены в числителе и знаменателе соответственно. Для наглядности приведены отличия характеристик, выраженные в процентах. Видно, что при малых нагрузках, характерных для многих реальных систем, расхождение измеренных величин достигает двух и более раз. Абсолютные значения характеристик меньше в системах с унимодальными потоками. Следовательно, могут возникать ошибки при проектировании реальных систем [15]. Для высоконагруженных систем значение расхождения имеет существенный разброс, однако в целом с увеличением загрузки системы погрешность замены полимодального распределения унимодальным уменьшается. Видно, что с увеличением загрузки существенно увеличивается доверительный интервал для измеренных значений (рис. 4). Это объясняется тем, что чем ближе система к перегруженному состоянию, тем значительно больше отклоняются измеряемые значения от средних. Также погрешность такой замены увеличивается с увеличением числа обслуживаемых приборов. Например, в системе с двумя приборами при нагрузке 0,9 погрешность средней длины очереди достигает 70%. На данную величину оказывает существенное влияние и межмодовый интервал. Характеристики систем со сравнительно малым межмодовым интервалом стремятся к характеристикам систем с унимодальным распределением. Это объясняется тем, что у системы с полимодальным распределением заявки склонны к группировке, и при большом межмодовом интервале для получения конкретного математического ожидания потребуется формировать группы из большего числа заявок. Чем больше приборов в системе массового обслуживания, тем более явно наблюдается эффект группировки заявок, следовательно, тем больше расхождение средних значений исследуемых характеристик.

Среднее время пребывания заявки в системе монотонно возрастает с увеличением загрузки системы ρ и межмодового интервала D . Коэффициент вариации двухмодового гиперэкспоненциального распределения зависит от D : уменьшается, начиная со значения $\nu_0 > 1$, равного коэффициенту вариации первой составляющей (в случае, когда составляющие распределения имеют одинаковый коэффициент вариации), принимая некоторое минимальное значение при незначительном влиянии мод друг на друга, и при дальнейшем увеличении D принимает постоянное значение $\nu < 1$. Для двухмодового гипоекспоненциального распределения коэффициент вариации также стремится к постоянному значению $\nu < 1$ (рис. 5). Выявленное свойство позволяет сделать вывод, что при большом межмодовом интервале для моделирования можно использовать простейший поток, поскольку его коэффициент вариации больше. Однако, как показали многочисленные эксперименты, если поток имеет полимодальное распределение,

то получаемые в ходе моделирования результаты существенно отличаются от реальных, а аппроксимация по математическому ожиданию и коэффициенту вариации не позволяет добиться адекватных результатов.

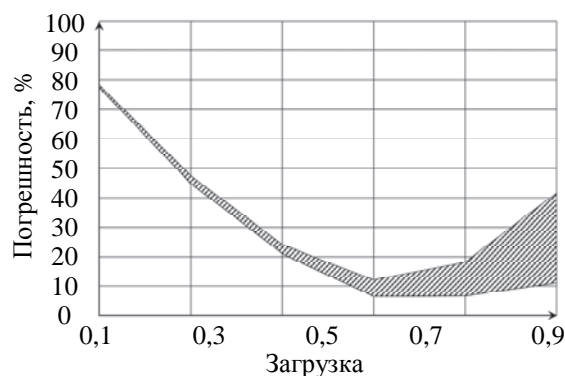


Рис. 4. Зависимость погрешности от загрузки системы

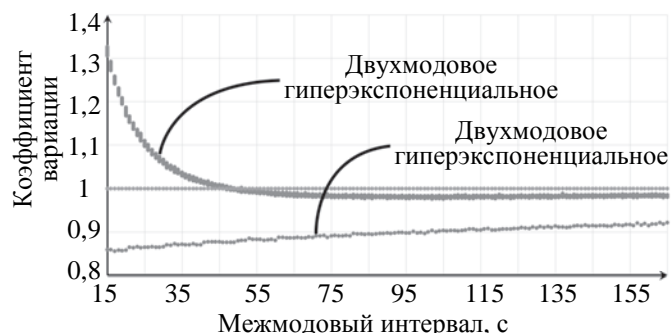


Рис. 5. Зависимость коэффициента вариации от межмодового интервала

Заключение

Обобщая рассмотренные выше свойства полимодальных распределений, можно сделать вывод, что для построения модели высоконагруженной системы с полимодальными потоками с допустимой инженерной погрешностью порядка 15% замена полимодального потока на унимодальный допустима. Следовательно, при моделировании высоконагруженных систем можно использовать унимодальные потоки вместо полимодальных. В частности, это существенно упрощает процесс проектирования высоконагруженных систем передачи данных, поскольку позволяет использовать классические методы моделирования, несмотря на полимодальность моделируемых потоков. С уменьшением загрузки системы увеличивается погрешность от замены полимодального распределения унимодальным. Для исследования систем с полимодальными потоками можно аппроксимировать реальные распределения совокупностью мультиэкспоненциальных, число мод которых равно числу мод в реальном распределении. Кроме того, при имитационном моделировании можно формировать поток заявок на основе гистограммы реального распределения или с помощью имеющейся трассы. Дальнейшие исследования вопроса могут предполагать комбинацию распределений с различающимися коэффициентами вариации и исследование комбинации из четырех и более мод.

Литература

1. Богатырев В.А., Богатырев С.В. Резервированная передача данных через агрегированные каналы в сети реального времени // Изв. вузов. Приборостроение. 2016. Т. 59. № 9. С. 735–740. doi: 10.17586/0021-3454-2016-59-9-735-740
2. Bogatyrev V.A. Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer // Automatic Control and Computer Sciences. 1999. V. 33. N 1. P. 57–63.
3. Aliev T.I., Rebezova M.I., Russ A.A. Statistical methods for monitoring travel agencies // Automatic Control and Computer Sciences. 2015. V. 49. N 6. P. 321–327. doi: 10.3103/S0146411615060024
4. Jasin A., Alsaqour R., Abdelhaq M., Alsukour O., Saeed R. Review on current transport layer protocols for TCP/IP model // International Journal of Digital Content Technology and its Applications. 2012. V. 6. N 14. P. 495–503. doi: 10.4156/jdcta.vol6.issue14.58

References

1. Bogatyrev V.A., Bogatyrev S.V. Redundant data transmission using aggregated channels in real-time network. *Journal of Instrument Engineering*, 2016, vol. 59, no. 9, pp. 735–740. (in Russian) doi: 10.17586/0021-3454-2016-59-9-735-740
2. Bogatyrev V.A. Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer. *Automatic Control and Computer Sciences*, 1999, vol. 33, no. 1, pp. 57–63.
3. Aliev T.I., Rebezova M.I., Russ A.A. Statistical methods for monitoring travel agencies. *Automatic Control and Computer Sciences*, 2015, vol. 49, no. 6, pp. 321–327. doi: 10.3103/S0146411615060024
4. Jasin A., Alsaqour R., Abdelhaq M., Alsukour O., Saeed R. Review on current transport layer protocols for TCP/IP model. *International Journal of Digital Content Technology and its Applications*, 2012, vol. 6, no. 14, pp. 495–503. doi: 10.4156/jdcta.vol6.issue14.58

5. Bogatyrev V.A., Parshutina S.A., Poptcova N.A., Bogatyrev A.V. Efficiency of redundant service with destruction of expired and irrelevant request copies in real-time clusters // *Communications in Computer and Information Science*. 2016. V. 678. P. 337–348. doi: 10.1007/978-3-319-51917-3_30
6. Bogatyrev V.A., Bogatyrev A.V. Functional reliability of a realtime redundant computational process in cluster architecture systems // *Automatic Control and Computer Sciences*. 2015. V. 49. N 1. P. 46–56. doi: 10.3103/S0146411615010022
7. Datey S.G., Ansari T. Mobile Ad-hoc networks its advantages and challenges // *International Journal of Electrical and Electronics Research*. 2015. V. 3. N 2. P. 491–496.
8. Богатырев В.А., Богатырев С.В. Эффективность резервирования и фрагментации пакетов при передаче по агрегированным каналам // *Изв. вузов. Приборостроение*. 2017. Т. 60. № 2. С. 165–170. doi: 10.17586/0021-3454-2017-60-2-165-170
9. Varshney L.R. Transporting information and energy simultaneously // *Proc. IEEE Int. Symp. on Information Theory, ISIT*. Toronto, Canada, 2008. P. 1612–1616. doi: 10.1109/ISIT.2008.4595260
10. Sridhar P., Sheikh-Bahaei S., Xia S., Jamshidi M. Multi-agent simulation using discrete event and soft-computing methodologies // *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*. Washington, 2003. V. 2. P. 1711–1716. doi: 10.1109/icsmc.2003.1244659
11. Ramesh Kumar K.R. Computer-aided design of MIC layout with postprocessor for photoplotter // *Defence Science Journal*. 1994. V. 44. N 4. P. 317–321. doi: 10.14429/dsj.44.4186
12. Prabhakar M., Singh J.N., Mahadevan G. Nash equilibrium and Markov chains to enhance game theoretic approach for vanet security // *Advances in Intelligent Systems and Computing*. 2013. V. 174. P. 191–199. doi: 10.1007/978-81-322-0740-5_24
13. Алиев Т.И. Аппроксимация вероятностных распределений в моделях массового обслуживания // *Научно-технический вестник информационных технологий, механики и оптики*. 2013. № 2 (84). С. 88–93.
14. Алиев Т.И. Трехмоментная аппроксимация вероятностных распределений в моделях массового обслуживания // *Научно-технический вестник информационных технологий, механики и оптики*. 2014. № 2 (90). С. 107–110.
15. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 7th ed. Springer, 2013. 745 p.
5. Bogatyrev V.A., Parshutina S.A., Poptcova N.A., Bogatyrev A.V. Efficiency of redundant service with destruction of expired and irrelevant request copies in real-time clusters. *Communications in Computer and Information Science*, 2016, vol. 678, pp. 337–348. doi: 10.1007/978-3-319-51917-3_30
6. Bogatyrev V.A., Bogatyrev A.V. Functional reliability of a real-time redundant computational process in cluster architecture systems. *Automatic Control and Computer Sciences*, 2015, vol. 49, no. 1, pp. 46–56. doi: 10.3103/S0146411615010022
7. Datey S.G., Ansari T. Mobile Ad-hoc networks its advantages and challenges. *International Journal of Electrical and Electronics Research*, 2015, vol. 3, no. 2, pp. 491–496.
8. Bogatyrev V.A., Bogatyrev S.V. Effectiveness of redundancy and packet fragmentation in transmission via aggregated channels. *Journal of Instrument Engineering*, 2017, vol. 60, no. 2, pp. 165–170. (in Russian) doi: 10.17586/0021-3454-2017-60-2-165-170
9. Varshney L.R. Transporting information and energy simultaneously. *Proc. IEEE Int. Symp. on Information Theory, ISIT*. Toronto, Canada, 2008, pp. 1612–1616. doi: 10.1109/ISIT.2008.4595260
10. Sridhar P., Sheikh-Bahaei S., Xia S., Jamshidi M. Multi-agent simulation using discrete event and soft-computing methodologies. *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*. Washington, 2003, vol. 2, pp. 1711–1716. doi: 10.1109/icsmc.2003.1244659
11. Ramesh Kumar K.R. Computer-aided design of MIC layout with postprocessor for photoplotter. *Defence Science Journal*, 1994, vol. 44, no. 4, pp. 317–321. doi: 10.14429/dsj.44.4186
12. Prabhakar M., Singh J.N., Mahadevan G. Nash equilibrium and Markov chains to enhance game theoretic approach for vanet security. *Advances in Intelligent Systems and Computing*, 2013, vol. 174, pp. 191–199. doi: 10.1007/978-81-322-0740-5_24
13. Aliev T.I. Approximation of probability distributions in queueing models. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2013, no. 2, pp. 88–93. (in Russian)
14. Aliev T.I. Three-moment based approximation of probability distributions in queueing systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2014, no. 2, pp. 107–110. (in Russian)
15. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 7th ed. Springer, 2013, 745 p.

Авторы

Жмылёв Сергей Александрович – ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0001-8916-5474, korg@cs.ifmo.ru

Алиев Тауфик Измайлович – доктор технических наук, профессор, заведующий кафедрой, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 57038212100, ORCID ID: 0000-0003-0278-688X, aliev@cs.ifmo.ru

Authors

Sergey A. Zhmylev – Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0001-8916-5474, korg@cs.ifmo.ru

Taufik I. Aliev – D.Sc., Professor, Head of Chair, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 57038212100, ORCID ID: 0000-0003-0278-688X, aliev@cs.ifmo.ru