

УДК 004.032
DOI: 10.15827/0236-235X.139.488-492

Дата подачи статьи: 26.07.22, после доработки: 28.07.22
2022. Т. 35. № 3. С. 488–492

Оценка граничных значений среднего времени отклика на запрос пользователя информационной системы

М.Н. Шелест¹, аспирант, *mshshelest@mail.ru*

Т.М. Татарникова¹, д.т.н., профессор, *tm-tatarn@yandex.ru*

¹ Санкт-Петербургский государственный университет аэрокосмического приборостроения, г. Санкт-Петербург, 190000, Россия

В работе предложены методы оценки верхней и нижней границ среднего времени отклика больших информационных систем на запрос пользователя.

Под запросом пользователя понимается транзакция, состоящая из последовательности команд, для выполнения которых резервируются вычислительные ресурсы. Транзакция формализуется в виде маршрута, представляющего собой цепочку пар «обслуживающее устройство–буфер», число которых равно числу команд транзакции. При этом обслуживающее устройство является средством выполнения команд транзакций, а буфер – памятью для фиксации результатов выполнения и ожидания времени для поступления на обслуживание. Выделение групп слабо связанных маршрутов позволяет выполнять параллельную обработку транзакций.

Предложена математическая схема большой информационной системы с группировкой маршрутов транзакций в виде сети массового обслуживания, так что каждая заявка пользователя проходит определенный маршрут из обслуживающих устройств. Метод оценки верхней границы среднего времени отклика системы на запрос пользователя основан на добавлении избыточных зависимостей и дублировании некоторых обслуживающих узлов, метод оценки нижней границы – на удалении обслуживающих узлов, играющих роль слабой связи между соседними маршрутами сети массового обслуживания.

Методы позволяют подобрать параметры, удовлетворяющие требованиям к разрабатываемой информационной системе и, соответственно, отвечающие показателям качества обслуживания пользователей информационных систем.

Ключевые слова: большая информационная система, сеть массового обслуживания, запрос пользователя, среднее время отклика на запрос, граничные значения среднего времени отклика.

Неотъемлемой частью полноценного функционирования многих организаций и предприятий являются большие *информационные системы* (ИС) [1, 2]. Для успешного запуска системы, которая будет удовлетворять заданным потребностям, еще на этапе проектирования необходимо проводить предварительный анализ характеристик ее функционирования [3, 4]. Одной из наиболее важных характеристик ИС является время отклика на запрос пользователя. Часто для оценки временных характеристик ИС прибегают к анализу модели системы, в роли которой выступает *сеть массового обслуживания* (СеМО). Время отклика ИС в данном случае соответствует среднему времени пребывания заявки в СеМО [5].

Однако даже в случае использования модели ИС вычисление истинного значения средней задержки для большинства СеМО является сложной, а порой и неразрешимой задачей из-за масштабности исследуемых систем. В таких ситуациях для оценки среднего времени пребывания заявок в сети прибегают к построению граничных значений [6, 7].

В данной работе рассмотрены методы построения верхнего и нижнего граничных значений для средней задержки заявки в СеМО.

Описание математической схемы большой ИС

Основной идеей нахождения граничных значений среднего времени пребывания заявки в СеМО является некоторое упрощение структуры исходной СеМО для возможности дальнейшего аналитического расчета.

Рассмотрим сеть, состоящую из пронумерованного набора элементарных систем массового обслуживания. Каждая из систем представляет собой пару «обслуживающее устройство–буфер», в которой обслуживающее устройство является средством выполнения команд транзакций, а буфер – памятью для фиксации результатов выполнения команд транзакций и ожидания времени для поступления на обслуживание. Предлагаемая модель ИС подразумевает, что для нее заранее известен набор последовательностей обслуживающих

устройств, в которых с заданной вероятностью будут обрабатываться транзакции. Таким образом, можно сказать, что каждая заявка пользователя ИС проходит определенный маршрут из обслуживающих устройств в СеМО. Также отметим, что при прохождении маршрута заявка не задерживается в промежуточных буферах, так как согласно правилу резервирования ресурсов для выполнения транзакций все обслуживающие устройства на маршруте заявки свободны [8]. Таким образом, набор маршрутов R представляется как список векторов с номерами обслуживающих устройств и полностью описывает структуру СеМО.

Каждый маршрут в СеМО характеризуется вероятностью появления заявок, обслуживаемых на маршруте, а каждое обслуживающее устройство – интенсивностью обслуживания и коэффициентом вариации времени обслуживания.

Методы определения верхней и нижней границ среднего времени отклика ИС на запрос

Предлагаемые в работе методы основаны на добавлении избыточных зависимостей и дублировании некоторых обслуживающих узлов при определении значения верхней границы быстродействия большой ИС и удалении обслуживающих узлов, играющих роль слабой связи между соседними маршрутами СеМО.

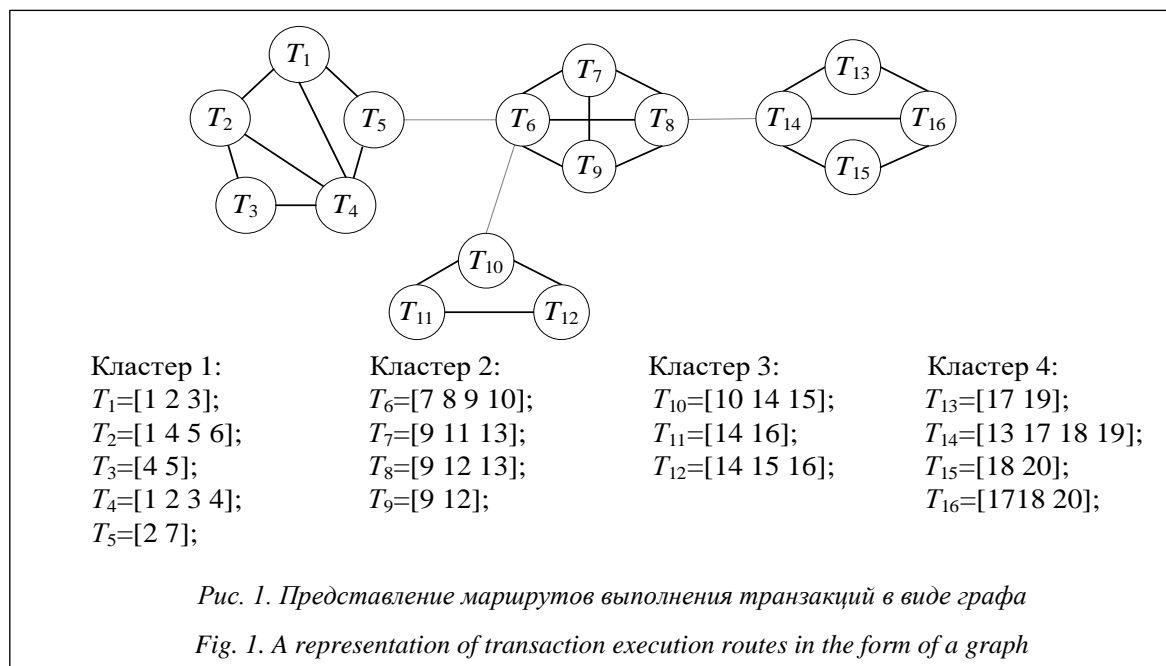
Рассмотрим модель большой ИС, в которой наблюдаются частые пересечения некоторых

маршрутов (рис. 1). Полное множество транзакций включает 16 маршрутов разной длины. Для обеспечения параллельной обработки запросов пользователей предлагается все множество маршрутов разбить на слабосвязанные группы маршрутов – кластеры.

Внутри каждого кластера одновременно может обслуживаться только одна заявка. Поступая на обслуживание, заявка блокирует все обслуживающие устройства своего маршрута. Необходимо понимать, что все смежные маршруты также окажутся заблокированными, поскольку имеют общие вычислительные ресурсы [9, 10]. Это означает, что за счет блокировки маршрутов, соединяющих соседние кластеры (то есть связующих узлов T_5, T_6, T_{10}, T_8 и T_{14} для графа на рис. 1), среднее время пребывания заявок в подграфах, образующих кластеры, не может быть рассчитано независимо от времени обработки заявок на подобных маршрутах.

Таким образом, предлагаются методы оценки верхнего и нижнего граничных значений среднего времени отклика ИС на запрос.

Метод оценки верхней границы подразумевает дублирование связующих узлов и введение в граф зависимости маршрутов дополнительных фиктивных соединений, чтобы кластер вместе со всеми смежными связующими узлами образовал отдельный полностью связанный подграф. Очевидно, что связующие узлы T_5, T_6, T_{10}, T_8 и T_{14} на рисунке 1 должны быть продублированы в каждом связанном с ними подграфе и к ним добавлены связи от остальных узлов



подграфа. За счет таких дополнений среднее время отклика ИС на запрос будет несколько увеличено, но зато позволит вычислить его для каждого кластера. Верхнее граничное значение среднего времени отклика ИС на запрос для всей СеМО вычисляется по следующей формуле:

$$\hat{T}^{UP_i} = \sum_{k=1}^K P_k \left[\frac{\Lambda_k}{2} \int_0^\infty t^2 \varphi_k(t) dt \int_0^\infty \frac{\varphi_k(t) dt}{\left(1 - \Lambda_k \int_0^t x \varphi_k(x) dx\right)^2} + \frac{1}{M_k} \right],$$

где P_k – вероятность поступления заявки в k -й полносвязный подграф; Λ_k – интенсивность поступления заявок в k -й полносвязный подграф; $\varphi_k(t)$ – плотность вероятности времени обслуживания в k -м полносвязном подграфе; M_k – интенсивность обслуживания заявок в k -м полносвязном подграфе.

Метод оценки нижней границы среднего времени отклика ИС на запрос заключается в поиске групп слабо связанных маршрутов, на которых не могут одновременно обрабатываться заявки. Такие группы представляют собой полносвязные подграфы, максимальные по включению вершин, – клики. Поиск клик производится на основе алгоритма Брона–Кербоша, после чего согласно алгоритму LBS (Load Based Strategy) из полученного набора клик выбираются отдельные клики, а связи между ними удаляются.

Согласно алгоритму LBS, разбиение на клики основано на величине загрузки множества маршрутов, входящих в состав клик. На каждом шаге для всех клик рассчитывается среднее время отклика ИС на запрос, после чего выбирается клика, для которой эта величина принимает наибольшее значение (рис. 2). В результате деления графа маршрутов на отдельные полносвязные подграфы с удалением соединяющих связей будет получен набор групп маршрутов. Все группы могут принимать заявки на обслуживание независимо друга от друга, но в каждой группе может обслуживаться только одна заявка. Таким образом, зная общую интенсивность обслуживания в каждой выделенной клике и плотность вероятности времен обслуживания в узлах подграфа, можно оценить среднее время пребывания заявки в преобразованной СеМО, состоящей из отдельных групп маршрутов.

На рисунке 3 приведены результаты моделирования работы СеМО и определения верхних и нижних граничных значений, найденных по предлагаемым методам.

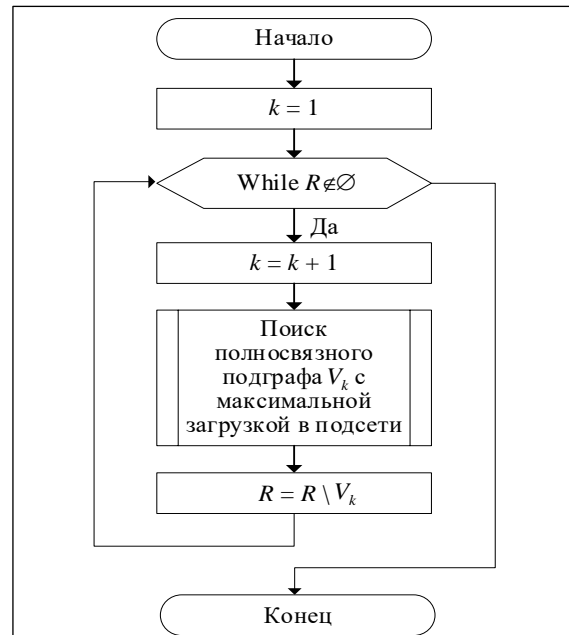


Рис. 2. Блок-схема алгоритма LBS

Fig. 2. A block diagram of the LBS algorithm

Отличительной особенностью методов нахождения верхнего граничного значения на основе избыточных зависимостей и нижнего граничного значения с помощью алгоритма LBS является высокая точность результатов, которую они показывают для больших ИС. Точность обусловлена тем, что большая ИС включает многочисленные ресурсы, которые сильно зависят друг от друга, и поэтому дублирование связующих узлов при нахождении верхней границы будет оказывать минимальное влияние на среднее время отклика ИС на запрос пользователя в каждом выделенном кластере. В свою очередь, выделение в отдель-

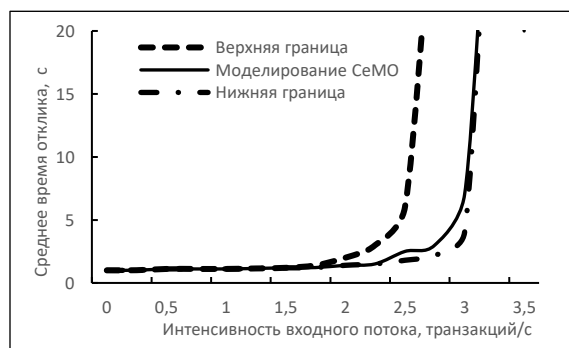


Рис. 3. Графики верхней и нижней границы для среднего времени отклика ИС на запрос

Fig. 3. Graphs of the upper and lower bounds for the information system average response time to a request

ные подсистемы более загруженных ресурсов для нахождения нижней границы приводит к беспрепятственной работе остальных частей системы.

Поскольку оба метода опираются на топологию графа маршрутов транзакций, для оценки влияния структуры СеМО на точность полученных оценок был проведен ряд экспериментов, анализ которых позволил сделать следующие выводы:

- точность верхней границы среднего времени пребывания заявки в СеМО напрямую зависит от размеров кластеров схожих транзакций;
- при довольно больших размерах кластеров методы дают возможность получить практически точную оценку временной характеристики системы;

– увеличение размеров кластеров приводит к сужению коридора возможных значений среднего времени пребывания заявки в СеМО.

Заключение

Анализ зависимостей верхней и нижней границ от интенсивности входного потока запросов пользователей позволяет не только оценить среднее время отклика ИС на запрос пользователя, но и выявить критическую точку перегрузки. Предложенные в работе методы дают возможность подобрать параметры ИС, удовлетворяющие требованиям к разрабатываемой системе и, соответственно, отвечающие показателям качества обслуживания пользователей ИС.

Литература

1. Проскуряков Н.Е., Ануфриева А.Ю. Анализ и перспективы современных систем хранения цифровых данных // Изв. ТулГУ: Технические науки. 2013. № 3. С. 368–377.
2. Gnanasundaram S., Shrivastava A. Information Storage and Management. NJ, John Wiley & Sons Publ., 2016, 544 p.
3. Фомин Д.С., Бальзамов А.В. Проблематика обработки транзакций при использовании микросервисной архитектуры. Изв. вузов. Поволжский регион. Технические науки. 2021. Т. 58. № 2. С. 15–23. DOI: 10.21685/2072-3059-2021-2-2.
4. Кутузов О.И., Татарникова Т.М. Инфокоммуникационные сети. Моделирование и оценка вероятностно-временных характеристик. СПб: Изд-во ГУАП, 2015. 382 с.
5. Shelest M.N., Bakin E.A. Analysis of parallel queueing network with mutual expectations. WECONF, 2018, pp. 1–4. DOI: 10.1109/WECONF.2018.8604361.
6. Kutuzov O.I., Tatarnikova T.M. On the acceleration of simulation modeling. Proc. XXII Int. Conf. SCM, 2019, pp. 45–47. DOI: 10.1109/SCM.2019.8903785.
7. Шелест М.Н. Анализ средней задержки для одной модели сети массового обслуживания с резервированием ресурсов // Информационно-управляющие системы. 2022. № 2. С. 32–41. DOI: 10.31799/1684-8853-2022-2-32-41.
8. Challawala S., Mehta C., Patel K., Lakhatariya J. MySQL 8 for Big Data: Effective Data Processing with MySQL 8, Hadoop, NoSQL APIs, and Other Big Data Tools. Packt Publishing, Birmingham, 2017, 266 p.
9. Богатырев В.А., Богатырев А.В., Богатырев С.В. Оценка надежности выполнения кластерами запросов реального времени // Изв. вузов. Приборостроение. 2014. Т. 57. № 4. С. 46–48.
10. Богатырев В.А., Кармановский Н.С., Попцова Н.А., Паршутина С.А., Воронина Д.А., Богатырев С.В. Имитационная модель поддержки проектирования инфокоммуникационных резервированных систем // Науч.-технич. вестн. информ. технологий, механики и оптики. 2016. Т. 16. № 5. С. 831–838. DOI: 10.17586/2226-1494-2016-16-5-831-838.

Boundary values evaluation for average response time to an information system user request

*M.N. Shelest*¹, Postgraduate Student, *mshshelest@mail.ru*

*T.M. Tatarnikova*¹, Dr.Sc. (Engineering), Professor, *tm-tatarn@yandex.ru*

¹ Saint Petersburg State University of Aerospace Instrumentation,
Saint Petersburg, 190000, Russian Federation

Abstract. The paper proposes methods for estimating the upper and lower bounds of the average response time of large information systems to a user request.

A user request is a transaction consisting of a sequence of commands for which computing resources are reserved. The transaction is formalized as a route, which is a chain of “server-buffer” pairs, their number is equal to the number of transaction commands. At the same time, the service device is a mean of executing transaction commands; the buffer is a memory for fixing the results of executing transaction commands and waiting for the time to arrive for service. Allocation of loosely coupled route groups allows parallel processing of transactions.

The authors propose a mathematical scheme of a large information system that organizes transaction routes in the form of a queuing network, so that each user request passes a certain route from the service devices. The method for estimating the upper bound on the average response time of the system to a user request is based on adding redundant dependencies and duplicating some service nodes. The method for estimating the lower bound of the average response time of the system to a user request is based on the removal of serving nodes that play the role of a weak connection between neighboring routes of the queuing network.

The proposed methods allow selecting such parameters that meet the requirements for the information system being developed and, accordingly, meet the indicators of the quality of service for users of information systems.

Keywords: big information system, queuing network, user request, average response time to a request, boundary values of the average response time.

References

1. Proskuryakov N.E., Anufrieva A.Yu. Analysis and prospects of modern systems of storage of figures. *Proceedings of the TSU*, 2013, no. 3, pp. 368–377 (in Russ.).
2. Gnanasundaram S., Shrivastava A. *Information Storage and Management*. NJ, John Wiley & Sons Publ., 2016, 544 p.
3. Fomin D.S., Balzamor A.V. The problem of transaction processing using microservice architecture. *University Proceedings. Volga Region. Technical Sciences*, 2021, vol. 58, no. 2, pp. 15–23. DOI: 10.21685/2072-3059-2021-2-2 (in Russ.).
4. Kutuzov O.I., Tatarnikova T.M. *Infocomm Networks. Modeling and Evaluation of Probabilistic-temporal Characteristics*. St. Petersburg, 2015, 382 p. (in Russ.).
5. Shelest M.N., Bakin E.A. Analysis of parallel queueing network with mutual expectations. *WECONF*, 2018, pp. 1–4. DOI: 10.1109/WECONF.2018.8604361.
6. Kutuzov O.I., Tatarnikova T.M. On the acceleration of simulation modeling. *Proc. XXII Int. Conf. SCM*, 2019, pp. 45–47. DOI: 10.1109/SCM.2019.8903785.
7. Shelest M.N. Average delay estimation for one queueing network model with resource reservation. *Information and Control Systems*, 2022, no. 2, pp. 32–41. DOI: 10.31799/1684-8853-2022-2-32-41 (in Russ.).
8. Challawala S., Mehta C., Patel K., Lakhataria J. *MySQL 8 for Big Data: Effective Data Processing with MySQL 8, Hadoop, NoSQL APIs, and Other Big Data Tools*. Packt Publishing, Birmingham, 2017, 266 p.
9. Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Estimation of reliability of execution of real-time queries. *J. of Instrument Engineering*, 2014, vol. 57, no. 4, pp. 46–48 (in Russ.).
10. Bogatyrev V.A., Karmanovskiy N.S., Poptsova N.A., Parshutina S.A., Voronina D.A., Bogatyrev S.V. Simulation model for design support of infocomm redundant systems. *Scientific and Technical J. of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 5, pp. 831–838. DOI: 10.17586/2226-1494-2016-16-5-831-838 (in Russ.).

Для цитирования

Шелест М.Н., Татарникова Т.М. Оценка граничных значений среднего времени отклика на запрос пользователя информационной системы // Программные продукты и системы. 2022. Т. 35. № 3. С. 488–492. DOI: 10.15827/0236-235X.139.488-492.

For citation

Shelest M.N., Tatarnikova T.M. Boundary values evaluation for average response time to an information system user request. *Software & Systems*, 2022, vol. 35, no. 3, pp. 488–492 (in Russ.). DOI: 10.15827/0236-235X.139.488-492.