В. А. БОГАТЫРЕВ
Университет ИТМО, Санкт-Петербург

А. В. БОГАТЫРЕВ, С. В. БОГАТЫРЕВ
АО Санкт-Петербургский центр компетенций NEO, Санкт-Петербург

РЕЗЕРВИРОВАННОЕ МНОГОПУТЕВОЕ ОБСЛУЖИВАНИЕ НЕОДНОРОДНОГО ПО КРИТИЧНОСТИ ЗАДЕРЖЕК ПОТОКА С ЗАДАНИЕМ ПУТЕЙ ПОСЛЕДОВАТЕЛЬНОГО ПРОХОЖДЕНИЯ УЗЛОВ ДЛЯ РЕПЛИК ЗАПРОСОВ

Предложена аналитическая модель оценки вероятности своевременности резервированного обслуживания неоднородного потока, предполагающего создание реплик запросов с кратностью, зависящей от суммарного допустимому времени ожидания в последовательности узлов, задействованных в обслуживании запроса.

Введение. Высокая надежность отказоустойчивость и готовность структурно-резервированных информационных систем и сетей [1–4] может поддерживаться на основе многопутевой маршрутизация [5, 6], при которой заранее прописывается основной и резервные маршруты (пути), что позволяет после отказов узлов основного достаточно быстро реконфигурировать систему с активацией резервного пути. Для инфокоммуникационных систем, работающих в реальном времени, в том числе в составе киберфизических систем [7–9], в ряде случаев для поддержки функциональной надежности требуется обеспечить непрерывность вычислительного процесса и своевременность обслуживания запросов при случайных отказах и злонамеренных воздействий [10–12].

Критерием эффективности многопутевого обслуживания в реальном времени является вероятность своевременного выполнения запросов с учетом прохождения через все узлы, последовательно задействованные в вычислительном процессе.

Вероятность своевременного и безошибочного обслуживания в кластерных системах и системах многопутевых передач в определенных условиях удается повысить в результате резервированного обслуживания, при котором создаются реплики запросов с заданием для каждой из них пути прохождения узлов, последовательно задействованных в вычислительном процессе (процессе передачи данных).

В исследуемых системах условием своевременности многопутевого резервированного обслуживания является прохождение хотя бы одной реплики всех узлов, составляющих путь ее обслуживания, за время меньшее заданного предельно допустимого значения.

Для неоднородного потока по критичности к времени ожидания запросов критерий эффективности заключается в вероятности своевременного выполнения всех типов запросов [7–9].

При неоднородности потока запросов по предельно допустимому времени ожидания, в данной статье исследуется потенциальная возможность повышения вероятности своевременного выполнения запросов всех типов в результате управления кратностью резервирования в зависимости от их критичности к задержкам в очередях.

Сложность оценки вероятности своевременности многопутевого резервированного обслуживания обусловлена необходимостью учета накопления времени ожидания в узлах, составляющих пути последовательного прохождения реплик запроса.

Особенность предлагаемой оценки искомой вероятности заключается вычисление при прохождении репликой очередного узла, включенного в маршрут ее обслуживания, вероятности не превышения остаточного допустимого времени ожидания с учетом средних задержек в очередях ранее пройденных узлов.

Предлагаемый доклад посвящен построению аналитической модели вероятности своевременности резервированного обслуживания неоднородного потока запросов и направлен на обоснование кратности резервирования запросов различной критичности к задержкам обслуживания с целью обеспечения своевременности обслуживания всех запросов неоднородного потока.

Модель резервированного обслуживания. Рассмотрим системы с неоднородностью входного потока по критичности к ожиданию запросов при их последовательном прохождении через m функциональных узлов. Будем предполагать, что i-й узел, зарезервирован с кратно-

стью n_i. Число создаваемых реплик (кратность резервирования) запросов зададим в зависимости от предельно допустимого времени их ожидания. Для каждой реплики задается маршрут (путь) последовательного прохождения m узлов, задействованных в ее обслуживании.

При прохождении одной из k_j реплик j-потока через i-й узел вероятность не превышения предельно-допустимого времени t_j с учет средних задержек в очередях ранее пройденных узлов вычислим как

$$P_{ij} = 1 - v_i \left[\sum_{j=1}^{N} \frac{\alpha_j \Lambda k_j}{n_i} \right] e^{\left(\frac{\alpha_j \Lambda k_j}{n_i} - \frac{1}{v_i} \right) \left(t_j - \sum_{g=1}^{i-1} w_g \right)},$$

где v_i – среднее время выполнения запроса i-м узлом с учетом всех N типов запросов, Λ – интенсивность суммарного потока запросов (без репликации), α_j – доля запросов j-го типа, t_j – предельно допустимое время ожидания для запросов j-го типа; wg – среднее время ожидания в g-м узле, включенном в путь выполнения [13] запроса,

$$w_g = \frac{\left(\Lambda_g v_g^2 \right)}{1 - \Lambda_g v_g},$$

$$\Lambda g = \sum_{j=1}^{N} \frac{\alpha_j \Lambda k_j}{n_g}.$$

Вероятность не превышения суммарного допустимого времени t_j при прохождении пути одной конкретной реплики запроса j-го типа и хотя бы одной из k_j реплик этого типа, определим соответственно как:

$$P_j = \prod_{i=1}^{m} P_{ij},$$

$$R_j = 1 - (1 - P_j)^{k_j}.$$

Вероятность своевременности обслуживания всех типов запросов неоднородного потока вычислим как:

$$R = \prod_{j=1}^{N} \left(1 - (1 - P_j)^{k_j} \right).$$

Результаты расчетов вероятности своевременного обслуживания. При расчетах будем предполагать, что имеется неоднородный поток, включающий два типа запросов (N=2), для первого из них предельно допустимое суммарное время ожидания во всех m узлах, задействованных в процессе обслуживания, t_1=0,1 с, а для второго t_2=0,5 с. Среднее время обслуживания запросов разных типов во всех узлах одинаково и равно v=0,1 с.

На рис. 1 приведена зависимость вероятности своевременного обслуживания запросов двух типов от интенсивности суммарного входного потока запросов Λ, а на рис. 2 от кратности резервирования запросов первого типа. Резервирование запросов второго типа не предполагается. На рис. 1 вероятности своевременного выполнения всех типов запросов R соответствуют кривые 1–3 при кратности резервирования запросов первого типа k= 1, 2, 5 и их доли α=0,1. На рис. 2 для интенсивности суммарного потока Λ=5 1/с кривые 1–4 соответствуют вероятности своевременного обслуживания запросов двух типов при α=0,2, 0,3, 0,4, 0,5. Расчет проведен при последовательном прохождении всех реплик запросов через m=5 узлов, каждый из которых зарезервирован с кратностью n=10/1.

Из представленных зависимостей можно заключить, что резервирование наиболее критичных к допустимому времени ожидания запросов позволяет повысить вероятность своевременного выполнения как наиболее критичных запросов, так и всей их совокупности. При этом существует оптимальная кратность резервирования запросов, зависящая от их критичности к допустимому времени ожидания, обеспечивающая максимум вероятности своевременного выполнения всех запросов неоднородного потока. Причем эффект от резервирования критичных к задержкам запросов наиболее рельефно проявляется в важном для практики случае обеспечения высокой вероятности своевременного выполнения запросов неоднородного потока.

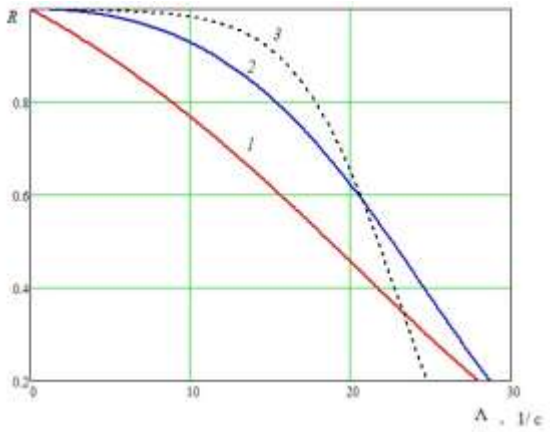

Рис. 1. Зависимость вероятности своевременного обслуживания от интенсивности входного потока запросов

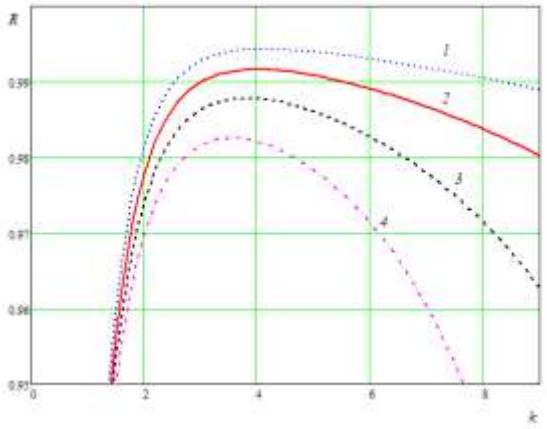

Рис. 2. Зависимость вероятности своевременного обслуживания запросов от кратности резервирования запросов первого типа

Предложенные модели могут быть использованы при оптимальном проектировании компьютерных систем реального времени, в том числе управляющих и коммуникационных систем, работающих [14–16] в составе киберфизических систем [17–19], разрабатываемых в рамках концепции построения высоконадежных систем с малыми задержками передач (Ultrareliable and Low-Latency Wireless Communication) [20–23].

Заключение. Предложена аналитическая модель оценки вероятности своевременности резервированного обслуживания, неоднородного потока, последовательностью резервированных узлов системы, предполагающей репликацию запросов с кратностью, зависящей от их критичности к допустимому времени ожидания запроса.

Показана, что резервирование наиболее критичных к допустимому времени ожидания запросов позволяет повысить вероятность своевременного выполнения всех запросов неоднородного потока.

Показано целесообразность оптимизации кратность резервирования запросов в зависимости от их критичности к допустимому времени ожидания.

ЛИТЕРАТУРА

1. Seontae Kim, Young-ri Choi. Constraint-aware VM placement in heterogeneous computing clusters. *Cluster Computing* 23(SI) March 2020 71-85.
2. Yang C., Liu J., Hsu C. *et al.* On improvement of cloud virtual machine availability with virtualization fault tolerance mechanism. *J Supercomput* 69, 1103–1122 (2014).
3. Jo C., Cho Y., Egger B.: A machine learning approach to live migration modeling. *In: Proceedings of the 2017 Symposium on Cloud Computing*, vol. 17, pp. 351–364. SoCC (2017)
4. Keller G., Lutfiyya H.: Dynamic management of applications with constraints in virtualized data centres. In: Proceedings of IFIP/IEEE International Symposium on Integrated Network Management (IM) (2015)
5. Prasenjit Chanak, Tuhina Samanta, Indrajit Banerjee Fault-tolerant multipath routing scheme for energy efficient wireless sensor networks. *International Journal of Wireless & Mobile Networks* (IJWMN) Vol. 5, No.2, April 2013 p 33-45.
6. Rajeev V., Muthukrishnan C.R. Reliable backup routing in fault tolerant real-time networks. Proceedings. *Ninth IEEE International Conference on Networks*, ICON 2001.
7. Bogatyrev, A.V., Bogatyrev, V.A., Bogatyrev, S.V. Multipath Redundant Transmission with Packet Segmentation (2019) *2019 Wave Electronics and its Application in Information and Telecommunication Systems, WECONF 2019*, art. no. 8840643. doi: 10.1109/WECONF.2019.8840643
8. Bogatyrev S.V., Bogatyrev V.A., Bogatyrev A.V. Redundant maintenance of a non-uniform query stream by a sequence of nodes that are grouped together in groups. *2020 Wave Electronics and its Application in Information and Telecommunication Systems*, WECONF 2020 9131463 DOI: 10.1109/WECONF48837.2020.9131463
9. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Timely Redundant Service of Requests by a Sequence of Cluster. *CEUR Workshop Proceedings*. 2020. Vol. 2590. pp. 1-12.
10. Yan J., Zhang M., Fu Z. An intralogistics-oriented Cyber-Physical System for workshop in the context of Industry 4.0 (2019) *Procedia Manufacturing*, 35, pp. 1178-1183. http://www.journals.elsevier.com/procedia-manufacturing doi: 10.1016/j.promfg.2019.06.074
11. Zakoldaev D.A., Korobeynikov A.G., Shukalov A.V., Zharinov I.O. Cyber and Physical Systems Technology Classification for Production Activity of the Industry 4.0 Smart Factory (2019) *IOP Conference Series: Materials Science and Engineering*, 582 (1), art. no. 012007. https://iopscience.iop.org/journal/1757-899X doi: 10.1088/1757-899X/582/1/012007

12. Bogatyrev V., Derkach A. Evaluation of a cyber-physical computing system with migration of virtual machines during continuous computing. Computers 2020 9(2),42doi:10.3390/computers9020042. www.mdpi.com/journal/computers
13. Kleinrock L. Queueing Systems: Volume I – Theory. New York: Wiley Interscience. 1975 p. 417. ISBN 978-0471491101
14. Sovetov B.Y., Tatarnikova T.M., Cehanovsky V.V. Detection system for threats of the presence of hazardous substance in the environment (2019) *Proceedings of 2019 22nd International Conference on Soft Computing and Measurements, SCM 2019*, art. no. 8903771, pp 121-124
15. Astakhova T.N., Verzun N.A., Kasatkin V.V., Kolbanev M.O., Shamin A.A. Sensor network connectivity models (2019) *Informatsionno-Upravliaiushchie Sistemy*, (5), pp. 38-50. http://www.i-us.ru/index.php/ius/article/view/4566/2609doi: 10.31799/1684-8853-2019-5-38-50
16. Ya S.B., Tatarnikova T.M., Poymanova E.D. Organization of multi-level data storage (2019) *Informatsionno-Upravliaiushchie Sistemy*, 2019 (2), pp. 68-75. https://www.i-us.ru/jour/article/view/469/399 doi: 10.31799/1684-8853-2019-2-68-75
17. **Абрамян Г.В**. Одноплатные компьютеры arduino как аппаратные средства программирования цифровых робототехнических и киберфизических систем // Преподавание информационных технологий в российской федерации: Материалы семнадцатой открытой всероссийской конференции /отв. ред. А.В. Альминдеров. 2019. с. 322-325.
18. Bogatyrev V.A. Fault Tolerance of Clusters Configurations with Direct Connection of Storage Devices. *Automatic Control and Computer Sciences - 2011*, Vol. 45, No. 6, pp. 330-337
19. Zakoldaev D.A., Korobeynikov A.G., Shukalov A.V., Zharinov I.O. Workstations Industry 4.0 for instrument manufacturing (2019) *IOP Conference Series: Materials Science and Engineering*, 665 (1), art. no. 012015. https://iopscience.iop.org/journal/1757-899X/ doi: 10.1088/1757-899X/665/1/012015
20. Murtaza Ahmed Siddiqi1, Heejung Yu and Jingon Joung. 5G Ultra-Reliable Low-Latency Communication Implementation Challenges and Operational Issues with IoT Devices. *Electronics* 2019, *8*, 981; doi:10.3390/electronics8090981 www.mdpi.com/journal/electronics
21. Ji H., Park S., Yeo J., Kim Y., Lee J., Shim B. Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects. *IEEE Wirel. Commun.* 2018, *25*, 124–130.
22. Sachs J., Wikström G., Dudda T., Baldemair R., Kittichokechai K. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. *IEEE Netw.* 2018, *32*, 24–31.
23. Bennis M., Debbah M., Poor H.V. Ultrareliable and Low-Latency Wireless Communication: Tail, Risk and Scale. *Proc. IEEE* 2018, *106*, 1834–1853.

V.A. Bogatyrev, (ITMO University, Saint Petersburg)
S.V. Bogatyrev, A.V. Bogatyrev JSC NEO Saint Petersburg Competence Center, Saint Petersburg)

Redundant Multipath Service of Non-uniform Criticality of Flow Delays with Setting Sequential Paths of Nodes for Query Replicas

Proposed analytical model estimates the probability of timely reserve service non-uniform flow, involving the creation of replicas queries with a frequency that depends on the total time allowed in the sequence of nodes involved in the service request