ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ / PHYSICAL & MATHEMATICAL SCIENCES

УДК 519.248 https://doi.org/10.33619/2414-2948/55/01

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРОЦЕССА ПОТРЕБЛЕНИЯ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ ПРИ ИХ НЕПОЛНОМ ОСВОБОЖДЕНИИ

*©**Носова М. Г.,** ORCID: 0000-0003-3641-7759, SPIN-код: 8091-3333, канд. физ.-мат. наук, Томский государственный университет систем управления и радиоэлектроники, г. Томск, Россия, nosovamgm@gmail.com*
*©**Садыков В. Д.,** Томский государственный университет систем управления и радиоэлектроники, г. Томск, Россия, sadykovvaris@hotmail.com*

MATHEMATICAL MODEL OF THE CONSUMPTION OF COMPUTER RESOURCES DURING THEIR INCOMPLETE RELEASE

*©**Nosova M.,** ORCID: 0000-0003-3641-7759, SPIN-code: 8091-3333, Ph.D., Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia, nosovamgm@gmail.com*
*©**Sadykov V.,** Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia, sadykovvaris@hotmail.com*

Аннотация. В статье разрабатывается математическая модель процесса потребления вычислительных ресурсов при их неполном освобождении при применении технологии виртуализации. Математическая модель представлена в виде системы массового обслуживания с неограниченным числом приборов, с простейшим входящим потоком заявок и экспоненциальным временем обслуживания их на приборах. Исследование модели выполнено методами теории массового обслуживания. Методом моментов найдены основные вероятностные характеристики величины объема свободных ресурсов: математическое ожидание и дисперсия. Предложенная математическая модель процесса потребления вычислительных ресурсов позволяет оценивать и прогнозировать процесс изменения объема свободных ресурсов виртуальной машины во времени и анализировать параметры ее производительности.

Abstract. The article develops a mathematical model of the process of consuming computing resources when they are incompletely released using virtualization technology. The mathematical model is presented in the form of a queuing system with an unlimited number of devices, with a simple incoming flow of applications and exponential service time for them on the devices. The study of the model is performed by the methods of queuing theory. Using the method of moments, the main probabilistic characteristics of the amount of free resources are found mathematical expectation and variance. The proposed mathematical model of the process of consuming computing resources allows us to evaluate and predict the process of changing the amount of free resources of a virtual machine in time and analyze its performance parameters.

Ключевые слова: система массового обслуживания, математическое моделирование, виртуальный сервер, оптимизация, виртуальная машина, потребление памяти.

Keywords: queuing system, mathematical modeling, virtual server, optimization, virtual machines, memory consumption.

Введение

Современное компьютерное оборудование обладает очень широкими возможностями, изначально оно было разработано для установки только одной операционной системы и для выполнения только одного приложения. Технология виртуализация смогла преодолеть это ограничение и сделала возможным одновременный запуск нескольких операционных систем и выполнение нескольких приложений на одном компьютере при этом такие виртуальные машины оказываются абсолютно изолированными друг от друга и ведут себя, как отдельные физические компьютеры. С помощью технологии виртуализации значительно повышается эффективность имеющегося в распоряжении предприятия оборудования за счет консолидации рабочих нагрузок, что позволяет существенно сократить количество серверного оборудования и снизить эксплуатационные издержки на содержание IT-инфраструктуры предприятия. На сегодняшний день миллионы людей во всем мире используют технологию виртуализации для экономии времени, денежных средств и электроэнергии, достигая при этом более высоких результатов без расширения аппаратных ресурсов [1].

Однако, пользователи виртуальных машин, могут столкнуться с такой проблемой, как утечка вычислительных ресурсов, которая возникает в результате неполного освобождения занятой памяти приложениями. Это приводит к тому, что потребление ресурсов в виртуальной машине неконтролируемо возрастает, в результате может наступить момент, когда новое выделение памяти для исполнения пользовательских запросов становится невозможным.

Поэтому с целью анализа и оптимизации параметров производительности виртуальной машины, актуальным является моделирование процесса изменения объема вычислительных ресурсов при их неполном освобождении при применении технологии виртуализации. Моделированию процесса потребления вычислительных ресурсов, посвящены многочисленные работы отечественных и зарубежных ученых. Значительная часть этих работ отводится применению при моделировании методов теории массового обслуживания [2–5]. Однако модели, разработанные учеными, не исчерпывают все их многообразие, актуальным по-прежнему остается разработка новых моделей, максимально приближенным к реальным процессам.

Математическая модель

В качестве новой математической модели процесса потребления виртуальной машиной вычислительных ресурсов при обработке запросов пользователей с неполным освобождением ресурсов выберем бесконечно линейную систему массового обслуживания M|M|∞ [6–7]. На вход такой системы M|M|∞ поступает простейший поток запросов с интенсивностью λ, время обслуживания запросов пользователей распределено по экспоненциальному закону с интенсивностью μ (Рисунок).

При поступлении заявки (запроса) в систему для выполнения выделяются вычислительные ресурсы, которые в большей степени освобождаются, когда запрос покидает систему. Если в момент поступления запроса недостаточно свободных вычислительных ресурсов виртуальной машины для его обработки, то такой запрос теряется. Будем считать, что потенциальный поток запросов к серверу является бесконечным.

Т. 6. №6. 2020
https://doi.org/10.33619/2414-2948/55

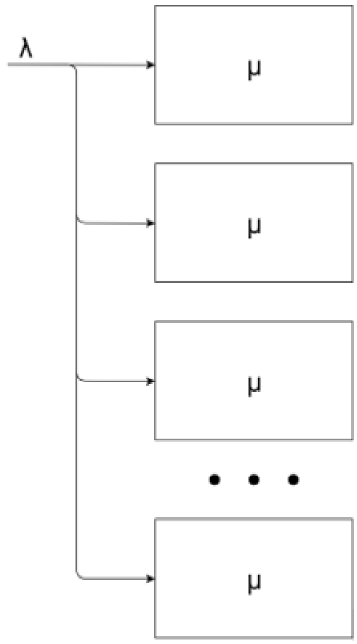

Рисунок. Система массового обслуживания M|M|∞.

Для функционирования данной системы важными являются два случайных процесса: число обсуживаемых запросов $k(t)$ в системе и объем свободных вычислительных ресурсов $S(t)$. Изменение числа обслуживаемых запросов $k(t)$ и объема свободных вычислительных ресурсов $S(t)$ происходит в следующих ситуациях:

1. На виртуальную машину поступает новый запрос. Полагая, что поток запросов на машину является простейшим с интенсивностью λ, запишем вероятность того, что за время Δt на виртуальную машину поступит новый запрос как $\lambda\Delta t + o(\Delta t)$. При этом каждый новый запрос связан с захватом некоторого объема ресурсов φ для его обработки. Размер φ является случайной величиной с функцией распределения $F_{\varphi}(x)$ и моментам $M\{\varphi\} = a_1$ и $M\{\varphi^2\} = a_2$.

2. В случайный момент времени запрос полностью обрабатывается виртуальной машиной. Будем полагать, что продолжительность обработки запроса является случайной величиной с функцией распределения $F(x) = 1 - e^{-\mu t}$. Каждый запрос обрабатывается и покидает виртуальную машину независимо от продолжительности обслуживания других запросов, с интенсивностью μ. Тогда за время Δt запрос покинет виртуальную машину с вероятностью $k\mu\Delta t + o(\Delta t)$. При этом каждый обработанный запрос частично освобождает ресурсы некоторого объема η. Размер η является случайной величиной с функцией распределения $F_{\eta}(x)$ и с моментами $M\{\eta\} = c_1$ и $M\{\eta^2\} = c_2$, причем, учитывая наличие утечки вычислительных ресурсов, всегда $c_1 < a_1$ и $c_2 < a_2$.

Нахождение вероятностных характеристик

В дальнейшем, применяя методы теории массового обслуживания [6–7], а также подход, описанный в [8–9], найдено, что изменение объема свободных вычислительный ресурсов при неполном освобождении ресурсов можно описать выражением: $V(t) = -\sum_{i=1}^{i(t)} \varphi + \sum_{l=1}^{l(t)} \eta$,

тогда общий объем свободных ресурсов виртуальной машины в момент времени t составит: $\tilde{V}(t) = V(0) + V(t)$, где $i(t)$ — случайный процесс, характеризующей число

Т. 6. №6. 2020
https://doi.org/10.33619/2414-2948/55

пришедших запросов за время t, а $l(t)$ — число запросов, завершающих обслуживание в момент времени t.

Тогда средняя величина изменения объема свободных ресурсов за время t составит:

$$MV(t) = \frac{\lambda}{\mu} c_1(1 - e^{-\mu t}) + \lambda a_1 t, M\tilde{V}(t) = V(0) + MV(t), \quad (1)$$

а дисперсия объема свободных ресурсов за время t:

$$\mathcal{D}\tilde{V}(t) = DV(0) + DV(t) = \quad (2)$$
$$= \frac{\lambda}{\mu} c_2(1 - e^{-\mu t}) - 2\frac{\lambda}{\mu} a_1 c_1(1 - e^{-\mu t}) - \lambda a_2 t.$$

где $V(0)$ — начальный объем свободных вычислительных ресурсов виртуальной машины.

Результатом исследования являются формулы (1) и (2), определяющие вероятностные характеристики величины объема свободных ресурсов виртуальной машины при их неполном освобождении, а именно математическое ожидание и дисперсию.

Заключение

Разработанная математическая модель процесса потребления вычислительных ресурсов и найденные вероятностные характеристики позволяет оценивать и прогнозировать процесс изменения объема свободных ресурсов виртуальной машины во времени и анализировать параметры ее производительности.

Список литературы:

1. Гусев О. В., Жуков А. В., Поляков В. В., Поляков С. В. Проблема адекватной оценки производительности веб-серверов в корпоративных сетях на предприятиях ЦБП // Новые информационные технологии в ЦБП и энергетике: материалы 6-й научно-технической конференции (Петрозаводск, 20-24 сентября 2004 г.). Петрозаводск, 2004. С. 84-87.

2. Vilaplana J., Solsona F., Teixidó I., Mateo J., Abella F., Rius J. A queuing theory model for cloud computing // The Journal of Supercomputing. 2014. V. 69. №1. P. 492-507. https://doi.org/10.1007/s11227-014-1177-y

3. Khazaei H., Misic J., Misic V. B. A fine-grained performance model of cloud computing centers // IEEE Transactions on parallel and distributed systems. 2012. V. 24. №11. P. 2138-2147. https://doi.org/10.1109/TPDS.2012.280

4. El-Sheimy N., Hou H., Niu X. Analysis and modeling of inertial sensors using Allan variance // IEEE Transactions on instrumentation and measurement. 2007. V. 57. №1. P. 140-149. https://doi.org/10.1109/TIM.2007.908635

5. Murugesan R., Elango C., Kannan S. A status report on resource allocation in cloud computing using queuing theory // International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 2014. V. 3. №11. P. 3603-3608.

6. Носова М. Г. Автономная немарковская система массового обслуживания и ее применение в задачах демографии: дисс. … канд. физ.-мат. наук. Томск, 2010. 204 с.

7. Назаров А. А., Терпугов А. Ф. Теория массового обслуживания. Томск, 2010. 228 с.

8. Носова М. Г. Математическая модель компании по микрофинансированию со смешанными потоками входящих рисков // Фундаментальные исследования. 2017. №12-1. С. 207-211.

9. Nosova M. G. Research of a three-phase autonomous queuing system with a Markov Modulated Poisson process // Information Technologies and Mathematical Modeling (ITMM-2018): Proceedings of 17th International Conference named after A. F. Terpugov (Tomsk, September 10-15, 2018). Tomsk, 2018. P. 33-38.

References:

1. Gusev, O. V., Zhukov, A. V., Polyakov, V. V., & Polyakov, S. V. (2004). Problema adekvatnoi otsenki proizvoditel’nosti veb-serverov v korporativnykh setyakh na predpriyatiyakh CBP. *Novye informatsionnye tekhnologii v CBP i energetike: materialy 6-i nauchno-tekhnicheskoi konferentsii (Petrozavodsk, 20-24 sentyabrya 2004 g.)*. Petrozavodsk, AO Metso-Automation, 2004, 84-87. (in Russian)

2. Vilaplana, J., Solsona, F., Teixidó, I., Mateo, J., Abella, F., & Rius, J. (2014). A queuing theory model for cloud computing. *The Journal of Supercomputing*, *69*(1), 492-507. https://doi.org/10.1007/s11227-014-1177-y

3. Khazaei, H., Misic, J., & Misic, V. B. (2012). A fine-grained performance model of cloud computing centers. *IEEE Transactions on parallel and distributed systems*, *24*(11), 2138-2147. https://doi.org/10.1109/TPDS.2012.280

4. El-Sheimy, N., Hou, H., & Niu, X. (2007). Analysis and modeling of inertial sensors using Allan variance. *IEEE Transactions on instrumentation and measurement*, *57*(1), 140-149. https://doi.org/10.1109/TIM.2007.908635

5. Murugesan, R., Elango, C., & Kannan, S. (2014). A status report on resource allocation in cloud computing using queuing theory. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *3*(11), 3603-3608.

6. Nosova, M. G. (2017). Matematicheskaya model’ kompanii po mikrofinansirovaniyu so smeshannymi potokami vkhodyashchikh riskov. *Fundamental’nye issledovaniya,* (12-1), 207-211. (in Russian)

7. Nazarov, A. A., & Terpugov, A. F. (2004). Teoriya massovogo obsluzhivaniya. Tomsk. (in Russian)

8. Nosova, M. G. (2017). A Mathematical model of microfinance company with mixed flows of incoming risks. *Fundamental’nye issledovaniya*, 12-1, 207-211. (in Russian).

9. Nosova, M. G. (2018). Research of a three-phase autonomous queuing system with a Markov Modulated Poisson process. *In Information Technologies and Mathematical Modeling (ITMM-2018): Proceedings of 17th International Conference named after A. F. Terpugov, September 10-15, 2018, Tomsk, Russia. Tomsk, 33-38.*

Работа поступила в редакцию 29.04.2020 г.

Принята к публикации 03.05.2020 г.

Ссылка для цитирования:

Носова М. Г., Садыков В. Д. Математическая модель процесса потребления вычислительных ресурсов при их неполном освобождении // Бюллетень науки и практики. 2020. Т. 6. №6. С. 10-14. https://doi.org/10.33619/2414-2948/55/01

Cite as (APA):

Nosova, M., & Sadykov, V. (2020). Mathematical Model of the Consumption of Computer Resources During Their Incomplete Release. *Bulletin of Science and Practice, 6*(6), 10-14. (in Russian). https://doi.org/10.33619/2414-2948/55/01