УДК 004. 75
DOI: 10.17586/0021-3454-2017-60-11-1033-1039

РЕЗЕРВИРОВАННОЕ ОБСЛУЖИВАНИЕ В ГРУППЕ ОДНОКАНАЛЬНЫХ СИСТЕМ С НАЗНАЧЕНИЕМ РАЗЛИЧНЫХ ПРИОРИТЕТОВ КОПИЯМ ЗАПРОСА

В. А. Богатырев[1], С. В. Богатырев[2]

[1]*Университет ИТМО, 197101, Санкт-Петербург, Россия*
E-mail: Vladimir.bogatyrev@gmail.com

[2]*Компания „Самсунг-электроникс“, Сеул, Корея*

Предложены дисциплина и модель резервированного обслуживания копий запросов в системе, узлы которой представляются одноканальными моделями массового обслуживания, отличающиеся тем, что для резервных копий запросов задаются разные приоритеты обслуживания. Запрос считается успешно обслуженным, если правильно выполнена хотя бы одна его копия. Исследованы возможности повышения эффективности систем резервированного обслуживания запросов в результате распределения приоритетов резервным копиям запроса. Показана эффективность дисциплины резервированного приоритетного обслуживания.

Ключевые слова: *кластер, дисциплина обслуживания, кратность резервирования, копии запросов, система массового обслуживания, приоритет*

Введение. К современным информационно-коммуникационным системам предъявляются высокие требования по производительности, отказоустойчивости и надежности функционирования при необходимости обеспечения безопасности [1—3] и непрерывной доступности (готовности) сервисов.

Для вычислительных систем ответственного назначения, эксплуатируемых в условиях сбоев, отказов и ошибок [4—8], высокая надежность и устойчивость функционировании достигается при резервировании как структуры, так и процессов передачи, хранения и обработки данных.

Резервированное выполнение запросов, поступающих при наличии незанятых каналов обслуживания, реализуется при дисциплине, названной в работах [9, 10] „широковещательное обслуживание с копированием запроса“. Такая дисциплина позволяет снизить среднее время ожидания запросов, однако не обеспечивает надежность обслуживания всех запросов, а не только поступающих при незанятости узлов обслуживания.

Организация и аналитические модели резервированного обслуживания всех запросов (вне зависимости от числа занятых каналов в момент их поступления) предложены и исследованы в работах [11—15].

Эффективность резервированного обслуживания для вычислительных систем кластерной архитектуры с реализацией очереди в каждом узле показана в [11—14], а для агрегированных каналов — в [15, 16]. Эффективность перераспределения запросов в мультикластерных системах исследована в работах [17, 18]. Возможности резервирования передач при многопутевой маршрутизации и распределении запросов через сеть по нескольким путям при поиске одного или нескольких серверов, выделяемых для выполнения копий запросов, проанализированы в работах [19, 20].

В дисциплинах резервированного обслуживания запросов, рассмотренных в [11—14], при поступлении каждого запроса создаются его копии, направляемые на выполнение

в разные узлы, моделируемые системами массового обслуживания (СМО) типа M/M/1 с бесконечными очередями.

Организация резервированного обслуживания связана с необходимостью разрешения технического противоречия, вызванного тем, что создание копий запросов, направляемых в разные узлы, повышая надежность вычислений (при необходимости получения безошибочного результата хотя бы для одной копии), вызывает возрастание загрузки узлов и, как следствие, увеличение задержек вычислений [11—15].

В то же время резервированное выполнение копий запросов разными узлами (повышая надежность вычислений), с учетом стохастичности обслуживания, может привести к сокращению ожидания как минимум в одном узле и тем самым повысить вероятность выполнения в требуемый срок хотя бы для одной копии запроса. При увеличении разброса времени ожидания в разных узлах потенциальная вероятность своевременного безошибочного выполнения хотя бы одной копии может быть увеличена. Срок ожидания первого результата выполнения копий запроса может быть сокращен с помощью предлагаемого подхода, предусматривающего использование разных дисциплин обслуживания копий, в том числе с варьированием приоритетов формируемых копий запросов.

В настоящей статье исследуются возможности повышения эффективности резервированного обслуживания запросов в случае назначения для копий запросов разных приоритетов.

Постановка задачи исследования. Рассмотрим вычислительную систему (кластер), объединяющую n компьютерных узлов, каждый из которых представим в виде одноканальной системы массового обслуживания. Модель исследуемой системы обслуживания приведена на рис. 1. При поступлении каждого запроса создаются k его копий, каждая из которых направляется в один из n ($k \le n$) узлов. В каждом узле организуется k очередей разного приоритета. Первая копия запроса направляется в очередь наибольшего приоритета одного из n узлов; вторая копия — в очередь второго приоритета другого узла и т д. Таким образом, первая копия имеет наивысший относительный приоритет, вторая — следующий по очереди приоритет и т д. Узлы, принимающие копии запросов для обслуживания, выбираются случайно или циклически. Обслуживание копий запросов в разных узлах будем считать независимым. Входной поток предположим простейшим, а время обслуживания — распределенным по показательному закону.

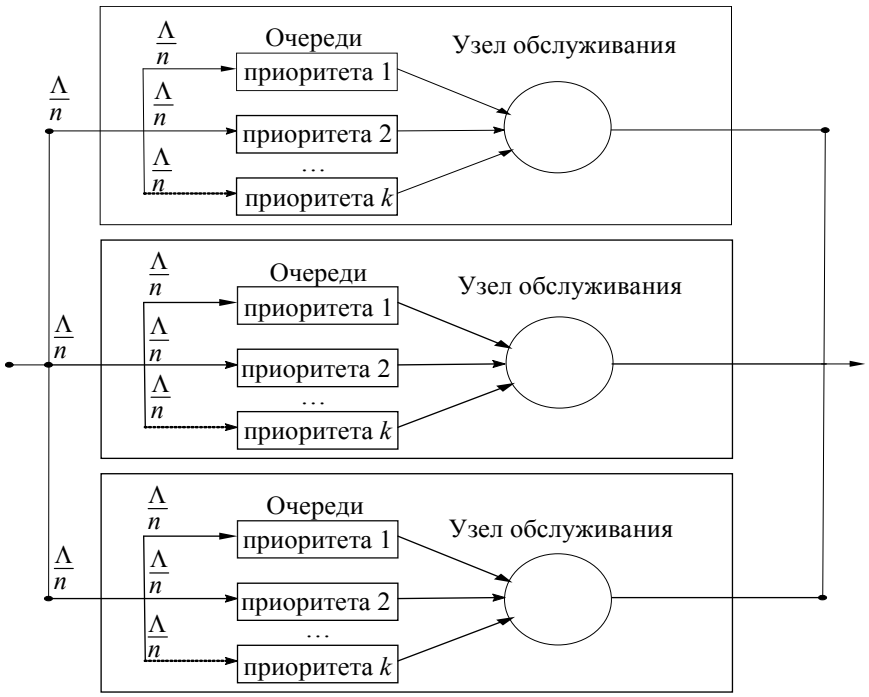

Рис. 1

При идентичности каналов вероятность правильного выполнения *i*-й копии запроса, имеющей *i*-й приоритет, при условии невыполнения или ошибочного выполнения всех предыдущих копий от первой до (*i*–1)-й вычисляется как $p(1-p)^{i-1}$, где p — вероятность безошибочного обслуживания копии запроса.

Эффективность A рассматриваемой дисциплины резервированного обслуживания определим как средний запас времени относительно предельно допустимой задержки ожидания w_0:

$$A = p\left(w_0 - w_1\right) + p(1-p)\left(w_0 - w_2\right) + p(1-p)^2\left(w_0 - w_3\right) + \ldots + p(1-p)^{k-1}\left(w_0 - w_k\right) = \\ = p\sum_{i=1}^{k}(1-p)^{i-1}\left(w_0 - w_i\right),$$

где w_i — среднее время ожидания для запросов *i*-го приоритета, i=1, 2, …, k.

При нормировании относительная эффективность F вычисляется как

$$F = \frac{p}{w_0}\sum_{i=1}^{k}(1-p)^{i-1}\left(w_0 - w_i\right).$$

При интенсивности входного потока Λ и создании k копий запроса (в случае балансировки загрузки узлов) в каждую из k очередей любого узла копии запросов поступают с интенсивностью $\Lambda_i = \Lambda/n$. Интенсивность потока запросов, поступающего на узел, при этом составляет $k\Lambda/n$.

При одинаковой длительности обслуживания всех копий $v_i = v$ и одинаковых их вторых начальных моментах $v_i^{(2)} = v^{(2)}$ среднее время ожидания копий запросов в очереди *i*-го приоритета определим как [21]

$$w_i = \left(\frac{k\Lambda v^{(2)}}{n}\right) \Big/ 2(1-r_{i-1})(1-r_i),$$

где для рассматриваемой дисциплины резервированного обслуживания $r_i = \Lambda v i / n$, $r_{i-1} = \dfrac{\Lambda v\left(i-1\right)}{n}$.

При экспоненциальном распределении длительности обслуживания имеем $v_i^{(2)} = 2v^2$ и соответственно

$$w_i = \left(\frac{k\Lambda v^2}{n}\right) \Big/ \left(1 - \frac{\Lambda v\left(i-1\right)}{n}\right)\left(1 - \frac{\Lambda v i}{n}\right).$$

С учетом условий стационарности режима обслуживания и ограничений на предельно допустимое время ожидания w_0 эффективность резервированного обслуживания с назначением приоритетов k копиям запросов определим как

$$F = \frac{p}{w_0}\sum_{i=1}^{k}(1-p)^{i-1}\left(w_0 - w_i\right)\delta_i,$$

где $\delta_i = 1$, если $\left[\left(\Lambda v i/n\right) < 1\right] \wedge \left[w_i \le w_0\right]$, иначе $\delta_i = 0$.

Вероятность безошибочного обслуживания копии запроса определяется исходя из реализации системы, в качестве узлов которой могут быть серверы кластера или каналы передачи данных.

При рассмотрении каналов передачи данных в качестве узлов системы вероятность безошибочности передачи пакета длиной L (бит) определяется как

$$p = (1 - B)^L,$$

где B — битовая вероятность ошибочной передачи, заметим, что при этом среднее время передачи пакета (выполнения запроса) длиной L можно определить как $v=L/s$, где s — скорость канала (бит/с).

Эффективность резервированного приоритетного обслуживания. Эффективность исследуемой дисциплины резервированного приоритетного обслуживания (с различной приоритетностью копий запросов) определим относительно дисциплины без резервированного обслуживания и дисциплины с резервированным бесприоритетным обслуживанием копий запросов.

Эффективность систем без резервированного обслуживания вычисляется как

$$F = \frac{p}{w_0}\left(w_0 - w_1\right) = \frac{p}{w_0}\left(w_0 - \frac{\left(\Lambda v^2/n\right)}{1-\left(\Lambda v/n\right)}\right)$$

при условии стационарности обслуживания, задаваемом как $\left(\Lambda v/n\right) < 1$.

Для систем с резервированным бесприоритетным обслуживанием копий запросов рассмотрим вариант реализации, при котором каналы обслуживания разделены на группы по k каналов и осуществляется одновременная выдача всех k копий запроса на выполнение во все каналы одной из групп. При таком варианте в результате резервированного обслуживания повышается вероятность безошибочного выполнения хотя бы одной копии запроса, при этом эффективность вычисляется как

$$F = \left(1-\left(1-p\right)^k\right)\frac{w_0 - w_1}{w_0} = \frac{\left(1-\left(1-p\right)^k\right)}{w_0}\left(w_0 - \frac{\left(k\Lambda v^2/n\right)}{1-\left(k\Lambda v/n\right)}\right)$$

при условии стационарности обслуживания $\left(k\Lambda v/n\right) < 1$.

Проведем расчет эффективности рассматриваемых дисциплин обслуживания в группе m=20 узлам (одноканальных СМО) для среднего времени выполнения запроса v =0,01 с и w_0=0,1 с. Зависимость эффективности обслуживания от интенсивности потока запросов представлена на рис. 2, *а*—*в*, где показаны результаты расчетов соответственно при вероятностях битовых ошибок B, равных $10^{-3,5}$, 10^{-4}, 10^{-5}. На рисунке кривые *1*, *2* соответствуют резервированному приоритетному обслуживанию при кратности копирования запросов k=3 и k=2; кривые *3*, *4* соответствуют вариантам резервированного бесприоритетного обслуживания при кратности копирования запросов k=3 и k=2, а кривая *5* — варианту нерезервированного обслуживания (k=1).

Анализ приведенных зависимостей позволяет констатировать преимущества предлагаемой приоритетной дисциплины резервированного обслуживания. Результаты расчетов показывают эффективность резервированного обслуживания при низкой интенсивности потока запросов (загрузки) и существование границы интенсивности, выше которой резервированное обслуживание становится нецелесообразным; при этом чем больше вероятность битовых ошибок, тем больше порог граничной интенсивности запросов и тем эффективнее резервированное выполнение запросов с более высокой кратностью. С увеличением вероятности битовых ошибок и уменьшением интенсивности потока запросов эффективность обслуживания

возрастает с ростом кратности резервирования, причем рост эффективности больше для приоритетной дисциплины резервированного обслуживания.

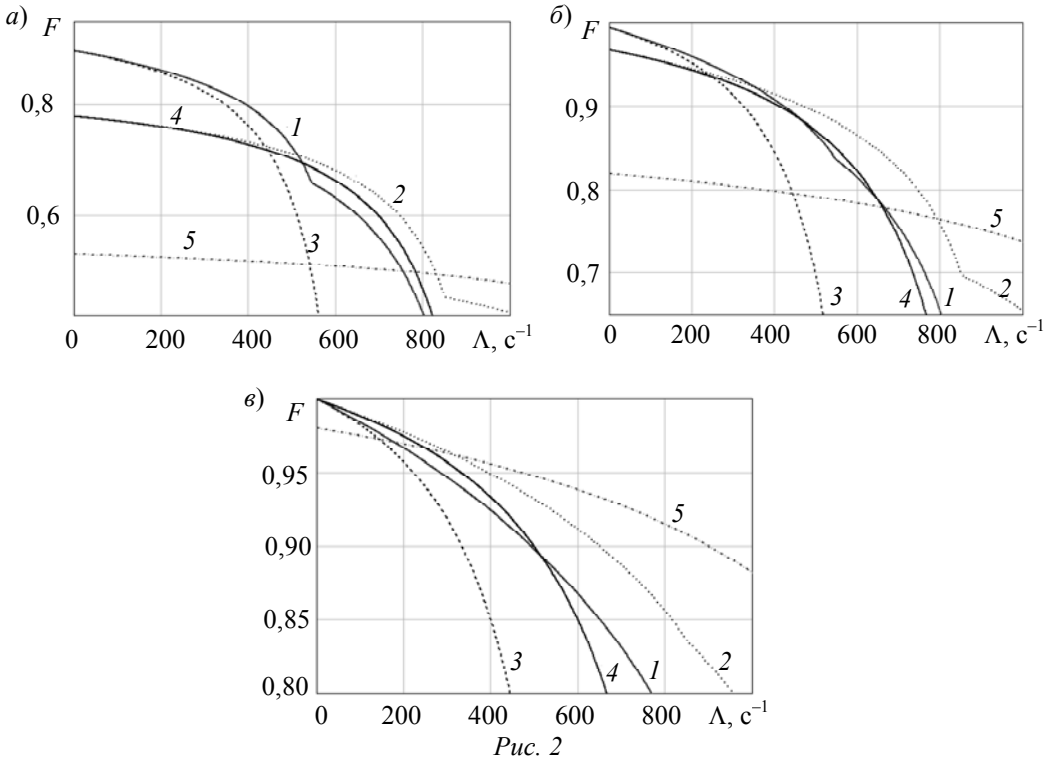

Рис. 2

Заключение. Предложены дисциплина и модель резервированного приоритетного обслуживания запросов в системе, узлы которой представляются одноканальными моделями массового обслуживания. Данные дисциплина и модель отличаются тем, что для резервных копий задаются разные приоритеты обслуживания, при этом запрос считается успешно обслуженным, если правильно выполнена хотя бы одна его копия.

Для систем резервированного обслуживания показаны преимущества предлагаемой приоритетной дисциплины относительно бесприоритетной в зависимости от интенсивности потока запросов и вероятности битовых ошибок.

Показана эффективность резервированного обслуживания и ее возрастание при увеличении кратности резервирования в соответствии с ростом вероятности битовых ошибок и снижении интенсивности потока запросов. Показано существование границы интенсивности запросов, ниже которой резервированное обслуживание целесообразно, причем пороговое значение граничной интенсивности увеличивается при увеличении вероятности битовых ошибок.

СПИСОК ЛИТЕРАТУРЫ

1. *Советов Б. Я., Колбанев М. О., Татарникова Т. М.* Технологии инфокоммуникации и их роль в обеспечении информационной безопасности // Геополитика и безопасность. 2014. № 1(25). С. 69—77.

2. *Гатчин Ю. А., Жаринов И. О., Коробейников А. Г.* Математические модели оценки инфраструктуры системы защиты информации на предприятии // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 2(78). С. 92—95.

3. *Алиев Т. И., Муравьева-Витковская Л. А.* Приоритетные стратегии управления трафиком в мультисервисных компьютерных сетях // Изв. вузов. Приборостроение. 2011. Т. 54, № 6. С. 44—48.

4. *Гуров С. В., Половко А. М.* Основы теории надежности. СПб: БХВ-Петербург, 2006. 704 с.

5. *Черкесов Г. Н.* Надежность аппаратно-программных комплексов. СПб: Питер, 2005. 479 с.

6. *Kopetz H.* Real-Time Systems: Design Principles for Distributed Embedded Applications. Springer, 2011. P. 396.

7. *Шубинский И. Б.* Функциональная надежность информационных систем: методы анализа // Надежность: науч.-техн. журн. 2012. 296 с.

8. *Богатырев, В. А.* Информационные системы и технологии. Теория надежности: Учеб. пособие. М.: Изд-во „Юрайт“, 2016. 318 с.

9. *Dudin A. N., Sun B.* A multiserver MAP/PH/N system with controlled broad-casting by unreliable servers // Automatic Control and Computer Sciences. 2009. N 43(5). P. 247—256.

10. *Lee M. H., Dudin A. N., Klimenok V. I.* The SM/V/N queueing system with broadcasting service // Math. Problems in Engineering. 2006. Vol. 2006. Art. ID 98171. 18 p.

11. *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, N 1. P. 46—56. DOI: 10.3103/S0146411615010022.

12. *Богатырев В. А., Богатырев А. В.* Модель резервированного обслуживания запросов реального времени в компьютерном кластере // Информационные технологии. 2016. Т. 22, № 5. С. 348—355.

13. *Богатырев В. А., Богатырев А. В.* Надежность функционирования кластерных систем реального времени с фрагментацией и резервированным обслуживанием запросов // Информационные технологии. 2016. Т. 22, № 6. С. 409—416.

14. *Богатырев В. А., Богатырев С. В.* Резервированное обслуживание в кластерах с уничтожением неактуальных запросов // Вестник компьютерных и информационных технологий. 2017. № 1(151). С. 21—28.

15. *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.

16. *Богатырев В. А., Сластихин И. А.* Эффективность резервированной передачи данных через агрегированные каналы // Изв. вузов. Приборостроение. 2016. Т. 59, № 5. С. 370—376.

17. *Богатырев В. А., Богатырев С. В.* Надежность мультикластерных систем с перераспределением потоков запросов // Изв. вузов. Приборостроение. 2017. Т. 60, № 2. С. 171—177.

18. *Богатырев В. А., Богатырев А. В., Богатырев С. В.* Перераспределение запросов между вычислительными кластерами при их деградации // Изв. вузов. Приборостроение. 2014. Т. 57, № 9. С. 54—58.

19. *Bogatyrev V. A., Parshutina S. A.* Redundant distribution of requests through the network by transferring them over multiple paths // Communications in Computer and Information Science, IET. 2016. Vol. 601. P. 199—207.

20. *Богатырев В. А., Паршутина С. А.* Модели многопутевой отказоустойчивой маршрутизации при распределении запросов через сеть // Вестник компьютерных и информационных технологий. 2015. № 12. С. 23—28.

21. *Алиев Т. И.* Основы моделирования дискретных систем: Учеб. пособие. СПб: СПбГУ ИТМО, 2009. 363 с.

Сведения об авторах

Владимир Анатольевич Богатырев — д-р техн. наук, профессор; Университет ИТМО; кафедра вычислительной техники; E-mail: Vladimir.bogatyrev@gmail.com

Станислав Владимирович Богатырев — Компания „Самсунг-электроникс“, Сеул; старший инженер; E-mail: realloc@gmail.com

Рекомендована кафедрой вычислительной техники

Поступила в редакцию 25.05.17 г.

Ссылка для цитирования: *Богатырев В. А., Богатырев С. В.* Резервированное обслуживание в группе одноканальных систем с назначением различных приоритетов копиям запроса // Изв. вузов. Приборостроение. 2017. Т. 60, № 11. С. 1033—1039.

REDUNDANT SERVICE IN GROUP OF SINGLE-CHANNEL SYSTEMS WITH DIFFERENT PRIORITIES ASSIGNED TO REQUEST COPIES

V. A. Bogatyrev[1], S. V. Bogatyrev[2]

[1]*ITMO University, 197101, St. Petersburg, Russia*
E-mail: Vladimir.bogatyrev@gmail.com
[2]*Samsung-Electronics, Seoul, Korea*

The discipline and a model of redundant service of request copies are proposed for a system with nodes represented by single-channel queuing model, in the case when different priorities of service are assigned to different backup requests. A request is considered handled successfully if at least one of its copies is served correctly. Possible ways to improve the efficiency of redundant service system by distributing priorities of backup requests. The efficiency of the proposed discipline of redundant priority service is demonstrated.

Keywords: cluster, queueing discipline, multiplicity of reservations, copies of requests, queueing system, priority

Data on authors

Vladimir A. Bogatyrev — Dr. Sci., Professor; ITMO University, Department of Computation Technologies; E-mail: Vladimir.bogatyrev@gmail.com

Stanislav V. Bogatyrev — Samsung-Electronics, Seoul; Senior Engineer; E-mail: realloc@gmail.com

For citation: Bogatyrev V. A., Bogatyrev S. V. Redundant service in group of single-channel systems with different priorities assigned to request copies. *Journal of Instrument Engineering*. 2017. Vol. 60, N 11. P. 1033—1039 (in Russian).

DOI: 10.17586/0021-3454-2017-60-11-1033-1039