

**Горбунова А.В.<sup>1</sup>, Самуйлов К.Е.<sup>1</sup>, Сопин Э.С.<sup>1,2</sup>**

<sup>1</sup> Российский университет дружбы народов, г. Москва, Россия

<sup>2</sup> Институт проблем информатики ФИЦ ИУ РАН, г. Москва, Россия

## **ПРЕОБРАЗОВАНИЕ ЛАПЛАСА-СТИЛТЬЕСА ДЛЯ ВРЕМЕНИ ОТКЛИКА СИСТЕМЫ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ С ГИСТЕРЕЗИСНЫМ УПРАВЛЕНИЕМ И ОГРАНИЧЕНИЕМ НА ОДНОВРЕМЕННОЕ ЧИСЛО АКТИВАЦИЙ**

### **АННОТАЦИЯ**

*Для анализа показателей производительности модели системы облачных вычислений с динамическим масштабированием в статье рассмотрена многолинейная система массового обслуживания с гистерезисным управлением количеством включенных приборов на основе порогов длины очереди и немедленным подключением дополнительных приборов, а также ограничением на одновременное количество активаций приборов с целью уменьшения вычислительной сложности без потери соответствующих свойств моделируемой системы. Разработан рекуррентный метод вычисления преобразования Лапласа-Стилтьеса распределения времени пребывания заявки в системе и времени ожидания начала обслуживания. С помощью полученного преобразования проведен анализ характеристик системы облачных вычислений.*

### **КЛЮЧЕВЫЕ СЛОВА**

*Облачные вычисления; время отклика; гистерезисное управление; динамическое масштабирование; пороговая система; преобразование Лапласа-Стилтьеса.*

**Gorbunova A.V.<sup>1</sup>, Samouylov K.E.<sup>1</sup>, Sopin E.S.<sup>1,2</sup>**

<sup>1</sup> RUDN University, Moscow, Russia

<sup>2</sup> Institute of Informatics Problems, FRC CSC RAS, Moscow, Russia

## **RESPONSE TIME ANALYSIS IN TERMS OF LAPLACE-STIELTJES TRANSFORM FOR CLOUD COMPUTING SYSTEM WITH HYSTERETIC CONTROL AND LIMITED SIMULTANEOUS ACTIVATIONS**

### **ABSTRACT**

*In the paper, cloud computing system with dynamic scaling is analyzed in terms of multiserver queuing system with hysteretic control of active server numbers based on queue length thresholds, and noninstantaneous server activation. To decrease computing complexity of the algorithms for evaluation of response time characteristics, we provide the analysis under simplifying assumption that only three servers may run activation procedures simultaneously. Recurrent algorithm for Laplace-Stieltjes transform of the response time and the sojourn time is developed. The algorithm is used to analyze response time characteristics of a cloud computing system.*

### **KEYWORDS**

*Cloud computing; response time; hysteretic control; dynamic scaling; Laplace-Stieltjes transform.*

### **Введение**

Под облачными вычислениями (англ. cloud computing) обычно понимается предоставление пользователю компьютерных ресурсов и мощностей в виде интернет-сервиса. Системы облачных вычислений применяются для хранения и обработки данных, для распределенных вычислений при решении научных и коммерческих задач. Современные облачные системы проектируются, как правило, масштабируемыми, что позволяет системе справляться с высокой нагрузкой и иметь возможность снизить энергопотребление в периоды снижения нагрузки.

Проблема недостаточного использования ресурсов возникает вследствие того, что их

распределение обычно происходит в соответствии с пиковой нагрузкой, которая может длиться только в течение короткого периода времени. Затем происходит переход в состояние простоя, но при этом потребляемая мощность может по-прежнему составлять около 50% от пиковой, что вызывает значительные потери энергии.

Вообще говоря, методы повышения энергетической эффективности и энергосбережения можно условно разделить на три категории: изменение рабочей нагрузки, влияние на поведение пользователей, перенастройка ресурсов. Идея первого заключается в использовании дополнительного устройства между сервисом и пользователем, которое изменяет интенсивность входящего трафика посредством буферизации: увеличивает продолжительность «спящего» режима и удлиняет периоды высокой загрузки [9]. Если же говорить о втором подходе, то положительный опыт внедрения мер стимулирования потребителей к энергосбережению имеется, в частности, у энергетических компаний. Применение системы тарификации, предполагающей значительное снижение стоимости электроэнергии в периоды низкой активности, например, ночью, побуждает потребителей к использованию электричества хотя бы отчасти в ночное время, что в свою очередь, приводит к выравниванию рабочей нагрузки. Однако у этих методов имеется существенный недостаток, который заключается в снижении качества оказываемых услуг, что является нарушением SLA.

Одним из способов реализации третьего подхода является динамическая активация виртуальных машин [5,10]. Иными словами, регулируется количество ресурсов (серверов, машин) в зависимости от текущей нагрузки [6-8]. При анализе таких систем применяются модели с пороговым управлением обслуживанием, в том числе, с гистерезисным управлением количеством включенных приборов [2,4]. При этом основной проблемой при моделировании облачных систем с помощью методов теории телетрафика является высокая вычислительная сложность получаемых в результате алгоритмов [1]. В [4] исследована система с гистерезисным управлением обслуживанием для анализа предоставления услуги видео по требованию, стационарные характеристики системы получены с помощью матричных методов, которые не применимы для анализа современных облачных платформ из-за высокой вычислительной сложности. В [10] система с гистерезисным управлением обслуживанием применена для анализа облачных систем с конечным числом приборов, а для случая трех приборов получен эффективный вычислительный алгоритм расчета стационарных характеристик функционирования облачной системы, имеющий линейную сложность. Однако слабой стороной предложенной модели является чрезвычайно большая мощность пространства состояний, которая имеет квадратичную зависимость от количества серверов. Поэтому в данной статье мы рассмотрим упрощенную модель с уменьшенным пространством состояний благодаря наложению ограничения на максимальное количество одновременно возможных активаций, и проведем анализ ее характеристик с помощью известных свойств преобразования Лапласа-Стилтьеса, алгоритм получения которого и будет предложен в данной работе. Статья является развитием работы [10], а также продолжением работы [3], в которой был разработан эффективный алгоритм вычисления стационарных вероятностей и показателей качества функционирования системы.

### **Математическая модель**

Рассматривается система облачных вычислений с гистерезисным подключением и отключением дополнительных виртуальных машин в виде многолинейной системы массового обслуживания с  $K$  приборами, часть которых может быть не активна, и конечной емкостью системы  $R$ . В систему поступает пуассоновский поток заявок с параметром  $\lambda$ . Считаем, что приборы являются однородными, время обслуживания распределено по экспоненциальному закону с параметром  $\mu$ .

В пустой системе активным, т.е. готовым при поступлении заявки мгновенно начать ее обслуживание, является один прибор. При поступлении заявок в систему активация приборов происходит не мгновенно, при этом количество активных приборов определяется числом заявок в очереди, в которой установлены парные пороги, заданные значениями векторов  $\mathbf{H} = (H_1, H_2, \dots, H_{K-1})$   $H_1 < H_2 < \dots < H_{K-1}$ , и  $\mathbf{L} = (L_1, L_2, \dots, L_{K-1})$ ,  $L_1 < L_2 < \dots < L_{K-1}$ , где  $L_{i+1} < H_i$ ,  $i = \overline{1, K-2}$  и  $L_i < H_i$ ,  $i = \overline{1, K-1}$ . Заявки обслуживаются в порядке поступления, т.е. очередь имеет дисциплину FCFS (First Come First Served). При поступлении на прибор заявка сохраняет место в очереди. Кроме того, накладывается ограничение на число одновременных активаций, которых не должно быть больше трех. Таким образом, когда в системе уже запущено два процесса активации, то при возникновении необходимости в запуске третьего дополнительного прибора его активация происходит мгновенно.

Правила работы системы следующие:

1. Если в системе уже есть  $H_i$  заявок, то при поступлении новой заявки активируется (подключается) один  $(i+1)$ -й дополнительный прибор, но не мгновенно, а через случайное время, имеющее экспоненциальное распределение с параметром  $\alpha$ ;
2. Если в системе находится  $L_i$  заявок и при этом одна заявка обслужилась, то  $(i+1)$ -й прибор мгновенно отключается, либо, если он не был активен, останавливается процедура его активации.

Функционирование системы описывается Марковским процессом  $X(t)$  с множеством состояний

$$S = \left\{ (k, i, n) \left| \begin{array}{l} 0 \leq n \leq H_1, k = 1, i = 1 \\ L_{k-1} \leq n \leq H_k, k = 2, i = \overline{k-1, k} \\ L_{k-1} \leq n \leq H_k, k = 3, K-1, i = \overline{k-2, k} \\ L_{k-1} \leq n \leq R, k = K, i = \overline{k-2, k} \end{array} \right. \right\},$$

где  $k$  – необходимое количество приборов;  $i$  – количество активированных приборов;  $n$  – количество заявок в очереди. Для выбранного расположения пороговых значений относительно друг друга диаграмма интенсивностей переходов для системы с  $K = 5$  приборами представлена на рисунке 1.

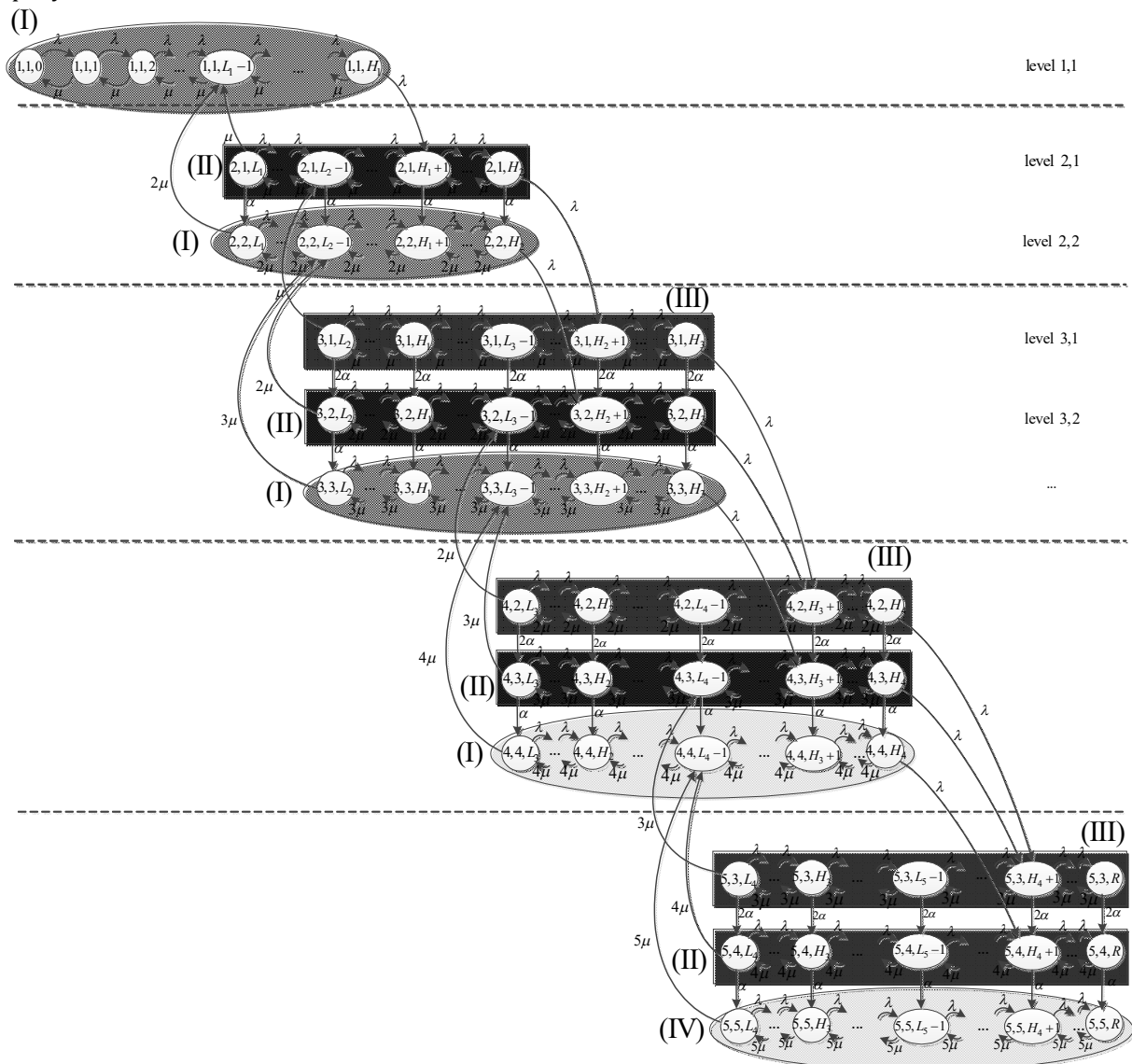


Рис. 1. Диаграмма интенсивностей переходов для системы с  $K=5$  приборами

## Метод вычисления преобразования Лапласа-Стилтьеса для времени отклика системы

Для того чтобы составить объективное мнение о поведении времени отклика системы, недостаточно знать только среднее значение этой случайной величины. Для всей полноты картины, а также для решения актуальных практических задач наряду с математическим ожиданием желательно иметь представление о значениях дисперсии и моментов высших порядков. В этой связи определение функции распределения времени отклика – в нашем случае в терминах преобразования Лапласа-Стилтьеса (ПЛС) – приобретает особую важность.

Обозначим  $V_{k,i,n}^m(s)$  – ПЛС времени ожидания  $m$ -й в очереди заявки, если система находится в состоянии  $(k,i,n)$ . В случае, когда  $m \leq i$ , время ожидания равно нулю, а значит

$$V_{k,i,n}^m(s) = 1, \quad m \leq i. \quad (1)$$

В остальных случаях будут справедливы рекуррентные соотношения, полученные, исходя из следующих соображений: время ожидания  $m$ -й в очереди заявки, если система находится в состоянии  $(k,i,n)$ , складывается из суммы двух времен, а именно времени, проведенного в ожидании перехода в одно из следующих возможных состояний и времени пребывания заявки в очереди в ожидании обслуживания, оставшегося уже после совершения этого шага, проделав который, естественно, что заявка может сменить свое местоположение в очереди. Теперь, воспользовавшись формулой полной вероятности, запишем следующие соотношения для различных состояний системы:

$$V_{k,k,n}^m(s) = \frac{k\mu}{\lambda + k\mu + s} V_{k,k,n-1}^{m-1}(s) + \frac{\lambda}{\lambda + k\mu + s} V_{k,k,n+1}^m(s), \quad (2)$$

$$1 \leq k \leq K, L_{k-1} < n < H_k;$$

$$V_{k,k,L_{k-1}}^m(s) = \frac{k\mu}{\lambda + k\mu + s} V_{k-1,k-1,L_{k-1}-1}^{m-1}(s) + \frac{\lambda}{\lambda + k\mu + s} V_{k,k,L_{k-1}+1}^m(s), \quad 2 \leq k \leq K; \quad (3)$$

$$V_{k,k,H_k}^m(s) = \frac{k\mu}{\lambda + k\mu + s} V_{k,k,H_k-1}^{m-1}(s) + \frac{\lambda}{\lambda + k\mu + s} V_{k+1,k,H_k+1}^m(s), \quad 1 \leq k \leq K-1; \quad (4)$$

$$V_{K,K,R}^m(s) = \frac{K\mu}{K\mu + s} V_{K,K,R-1}^{m-1}(s); \quad (5)$$

$$V_{k,k-2,n}^m(s) = \frac{(k-2)\mu}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-2,n-1}^{m-1}(s) + \frac{\lambda}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-2,n+1}^m(s) + \frac{2\alpha}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-1,n}^m(s), \quad 2 \leq k \leq K, L_{k-1} < n < H_k; \quad (6)$$

$$V_{k,k-1,n}^m(s) = \frac{(k-1)\mu}{\lambda + (k-1)\mu + \alpha + s} V_{k,k-1,n-1}^{m-1}(s) + \frac{\lambda}{\lambda + (k-1)\mu + \alpha + s} V_{k,k-1,n+1}^m(s) + \frac{\alpha}{\lambda + (k-1)\mu + \alpha + s} V_{k,k,n}^m(s), \quad 2 \leq k \leq K, L_{k-1} < n < H_k; \quad (7)$$

$$V_{k,k-2,L_{k-1}}^m(s) = \frac{(k-2)\mu}{\lambda + (k-2)\mu + 2\alpha + s} V_{k-1,k-2,L_{k-1}-1}^{m-1}(s) + \frac{\lambda}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-2,L_{k-1}+1}^m(s) + \frac{2\alpha}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-1,L_{k-1}}^m(s), \quad 2 \leq k \leq K; \quad (8)$$

$$V_{k,k-1,L_{k-1}}^m(s) = \frac{(k-1)\mu}{\lambda + (k-1)\mu + \alpha + s} V_{k-1,k-1,L_{k-1}-1}^{m-1}(s) + \frac{\lambda}{\lambda + (k-1)\mu + \alpha + s} V_{k,k-1,L_{k-1}+1}^m(s) + \frac{\alpha}{\lambda + (k-1)\mu + \alpha + s} V_{k,k,L_{k-1}}^m(s), \quad 2 \leq k \leq K; \quad (9)$$

$$V_{k,k-2,H_k}^m(s) = \frac{(k-2)\mu}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-2,H_k-1}^{m-1}(s) + \frac{\lambda}{\lambda + (k-2)\mu + 2\alpha + s} V_{k+1,k-1,H_k+1}^m(s) + \frac{2\alpha}{\lambda + (k-2)\mu + 2\alpha + s} V_{k,k-1,H_k}^m(s), \quad 2 \leq k \leq K-1; \quad (10)$$

$$V_{k,k-1,H_k}^m(s) = \frac{(k-1)\mu}{\lambda + (k-1)\mu + \alpha + s} V_{k,k-1,H_{k-1}}^{m-1}(s) + \frac{\lambda}{\lambda + (k-1)\mu + \alpha + s} V_{k+1,k-1,H_{k+1}}^m(s) + \frac{\alpha}{\lambda + (k-1)\mu + \alpha + s} V_{k,k,H_k}^m(s), \quad 2 \leq k \leq K-1; \quad (11)$$

$$V_{K,i,R}^m(s) = \frac{i\mu}{i\mu + (K-i)\alpha + s} V_{K,i,R-1}^{m-1}(s) + \frac{(K-i)\alpha}{i\mu + (K-i)\alpha + s} V_{K,i+1,R}^m(s), \quad 1 \leq i < K. \quad (12)$$

Далее на основе рекуррентных соотношений (1)–(12) составляется алгоритм уже для вычисления ПЛС  $V_{k,i,n}^n(s)$ , с помощью которого можно найти ПЛС  $V(s)$  времени ожидания и ПЛС  $W(s)$  времени пребывания заявки в системе:

$$V(s) = \sum_{(k,i,n) \in S} \pi_{k,i,n} V_{k,i,n}^n(s); \quad (13)$$

$$W(s) = \frac{\mu}{\mu + s} V(s), \quad (14)$$

где  $\pi_{k,i,n}$  – стационарные вероятности для соответствующих состояний  $(k,i,n)$ , алгоритм вычисления которых приведен в [3]. Последовательность применения рекуррентных формул для вычисления  $V_{k,i,n}^n(s)$  в силу громоздкости ее описания приводить не будем, но заметим, что в качестве первых шагов итерации следует рассматривать ПЛС для времени ожидания обслуживания заявки, стоящей первой в очереди, постепенно отдавая ее от обслуживающих приборов.

Задача обращения ПЛС для времени ожидания начала обслуживания и времени отклика системы в случае большого объема накопителя и, соответственно, значений парных порогов активации и отключения дополнительных приборов, представляет собой значительную вычислительную трудность. Однако благодаря свойствам ПЛС, в частности, тому, что  $n$ -й момент случайной величины  $\xi$  равен:

$$E\xi^n = (-1)^n \pi^{(n)}(0),$$

где  $\pi^{(n)}(0)$  – это значение  $n$ -й производной ПЛС случайной величины  $\xi$  в нуле, возможно вычислить моменты высших порядков и составить полноценное представление о поведении исследуемых случайных величин.

### Численный пример

Рассмотрим применение разработанного метода расчета временных характеристик модели системы облачных вычислений в терминах преобразования Лапласа-Стилтьеса для следующего набора данных: число приборов  $K = 4$ , максимальная длина очереди  $R = 75$ , векторы нижних и верхних порогов  $L = (10,20,30)$ ,  $H = (25,35,45)$ , интенсивность обслуживания заявок  $\mu = 1$ .

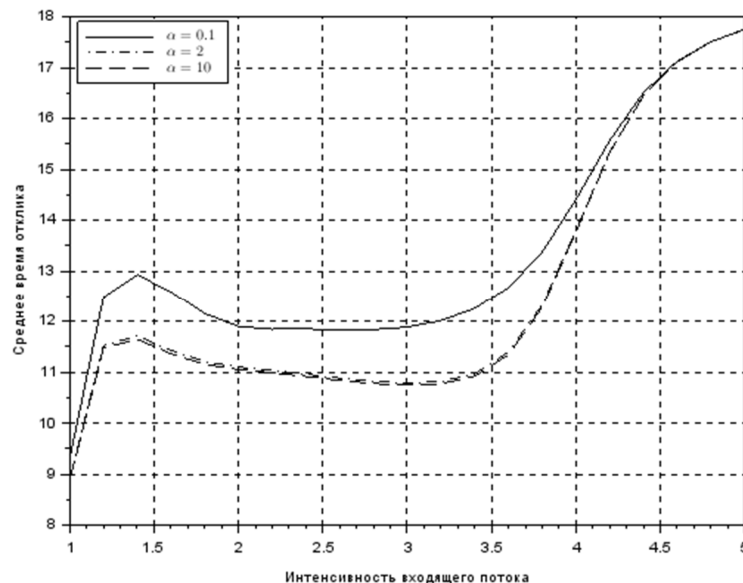


Рис.2. Математическое ожидание времени отклика

На рисунке 2 изображены графики зависимости среднего времени отклика от интенсивности входящего потока  $\lambda$  для различных значений параметра  $\alpha$  экспоненциального времени активации дополнительных приборов, на которых в целом наблюдается рост математического ожидания, что вполне естественно для увеличения значений интенсивности входящего потока. При  $\alpha = 0.1$  колебания среднего времени отклика незначительны, и с ростом  $\lambda$  кривая сглаживается, поскольку большую часть времени заявки проводят в ожидании начала обслуживания. Для двух других значений  $\alpha$  на графиках более четко выражены локальные минимумы, что объясняется выигрышем во времени при небольших значениях загрузки системы за счет быстрого и своевременного подключения дополнительных приборов.

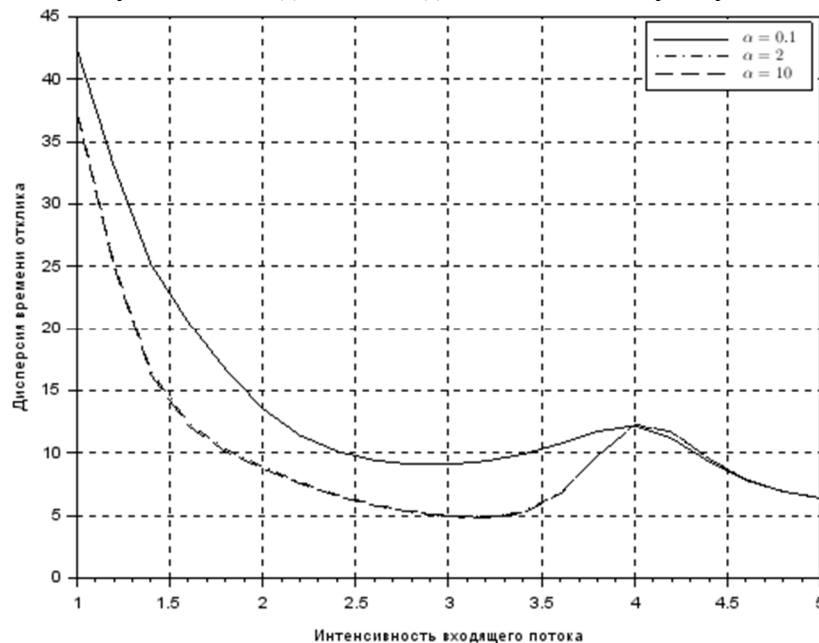


Рис.3. Дисперсия времени отклика

Как видно из рисунка 3, на графике дисперсии тоже присутствуют локальные минимумы, что также вызвано вышеописанной причиной. При малом значении  $\alpha = 0.1$  система ведет себя более определенно, так как серверы в преобладающем большинстве случаев не успевают вовремя подключиться и заявкам приходится длительную часть времени пребывать в очереди. При больших значениях  $\alpha$  наблюдаются скачки, в частности, локальный максимум при  $\lambda = 4$ .

### Заключение

Разработанный метод расчета временных характеристик модели системы облачных вычислений в терминах преобразования Лапласа-Стилтьеса в сочетании с полученными ранее результатами позволяет с высокой точностью получить оценки важнейших показателей качества обслуживания, на которые следует ориентироваться при расчете и проектировании облачных систем. Однако, для реальных систем облачных вычислений, где вычислительные облака включают тысячи серверов и обеспечивают ресурсами десятки тысяч приложений, которые одновременно используют миллионы пользователей, размерность пространства состояний модели создает вычислительные трудности, даже несмотря на линейную вычислительную сложность предложенного в статье алгоритма. Поэтому задачей дальнейших исследований является разработка приближенного метода оценки показателей производительности модели системы облачных вычислений.

*Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научных проектов № 14-07-00090, 15-07-03051, 15-07-03608.*

### **Литература**

1. Basharin G.P., Gaidamaka Yu.V., Samuilov K.E. Mathematical teletraffic theory and its application to the analysis of the next generations multiservice networks // Automatic control and computer sciences, Latvia, Riga: Institute of electronics and computer science of the Latvian university. - 2013. - № 2. - P. 11-21.
2. Гайдамака Ю.В., Печинкин А.В., Разумчик Р.В., Самуйлов А.К., Самуйлов К.Е., Соколов И.А., Сопин Э.С., Шоргин С.Я. Распределение времени выхода из множества состояний перегрузки в системе  $M | M | 1 | <L, N> | <H, R>$  с гистерезисным управлением нагрузкой // Информатика и ее применения. - 2013. - Т. 7, Вып. 4. - С. 20-33.

3. Gaidamaka Yu., Sopin E., Talanova M. A simplified model for performance analysis of cloud computing systems with dynamic scaling // Proc. of the 18th International Scientific Conference "Distributed Computer and Communication Networks: Control, Computation, Communications" DCCN-2015, M.: ИПУ РАН, 2015. - P. 75-86
4. Golubchik L., Lui J.C.S. Bounding of Performance Measures for Threshold-Based Queuing Systems: Theory and Application to Dynamic Resource Management in Video-on-Demand Servers // IEEE Trans. Computers. - 2002. - Vol. 51, №. 4. - P. 353-372.
5. Goswami V., Patra S.S., Mund G.B. Performance Analysis of Cloud with Queue-Dependent Virtual Machines // Proc. of 1st Int'l Conf. on Recent Advances in Information Technology, Dhanbad, India. - 2012. - P. 357-362.
6. Kaxiras S., Martonosi M. Computer Architecture Techniques for Power-Efficiency // Synthesis Lectures on Computer Architecture. - 2008. - Vol. 3, №. 1. - Pp. 1-207.
7. Lin M., Wierman A., Andrew L.L.H., Thereska E. Dynamic right-sizing for power-proportional data centers // INFOCOM, Proceedings IEEE. - 2011. - P. 1098-1106.
8. Meisner D., Gold B.T., Wenisch T.F. PowerNap: eliminating server idle power // CM SIGPLAN Notices. - 2009. - Vol. 44. - P. 205-216.
9. Nedeveschi S., Popa L., Iannaccone G., Ratnasamy S., Wetherall D. Reducing network energy consumption via sleeping and rate-adaptation // Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, NSDI'08, Berkeley, CA, USA. USENIX Association. - 2008. - P. 323-336.
10. Shorgin S.Y., Pechinkin A.V., Samouylov K.E., Gaidamaka Y.V., Gudkova I.A., Sopin E.S. Threshold-based Queuing System for Performance Analysis of Cloud Computing System with Dynamic Scaling // Proc. of the 12th International Conference of Numerical Analysis and Applied Mathematics ICNAAM-2014, Rhodes, Greece, 2014, USA, AIP Publishing - 2015. - Vol. 1648. - P. 1-3.

## References

1. Basharin G.P., Gaidamaka Yu.V., Samuilov K.E. Mathematical teletraffic theory and its application to the analysis of the next generations multiservice networks // Automatic control and computer sciences, Latvia, Riga: Institute of electronics and computer science of the Latvian university. - 2013. - №. 2. - P. 11-21.
2. Gaydamaka Yu.V., Pechinkin A.V., Razumchik R.V., Samuylov A.K., Samuylov K.E., Sokolov I.A., Sopin E.S., Shorgin S.Ya. Raspreделение vremeni vykhoda iz mnozhestva sostoyaniy peregruzki v sisteme  $M | M | 1 | <L, H> | <H, R>$  s gisterезisnym upravleniem nagruzkoй // Informatika i ee primeneniya. - 2013. - T. 7, Vyp. 4. - S. 20-33.
3. Gaidamaka Yu., Sopin E., Talanova M. A simplified model for performance analysis of cloud computing systems with dynamic scaling // Proc. of the 18th International Scientific Conference "Distributed Computer and Communication Networks: Control, Computation, Communications" DCCN-2015, M.: ИПУ РАН, 2015. - P. 75-86
4. Golubchik L., Lui J.C.S. Bounding of Performance Measures for Threshold-Based Queuing Systems: Theory and Application to Dynamic Resource Management in Video-on-Demand Servers // IEEE Trans. Computers. - 2002. - Vol. 51, №. 4. - P. 353-372.
5. Goswami V., Patra S.S., Mund G.B. Performance Analysis of Cloud with Queue-Dependent Virtual Machines // Proc. of 1st Int'l Conf. on Recent Advances in Information Technology, Dhanbad, India. - 2012. - P. 357-362.
6. Kaxiras S., Martonosi M. Computer Architecture Techniques for Power-Efficiency // Synthesis Lectures on Computer Architecture. - 2008. - Vol. 3, №. 1. - Pp. 1-207.
7. Lin M., Wierman A., Andrew L.L.H., Thereska E. Dynamic right-sizing for power-proportional data centers // INFOCOM, Proceedings IEEE. - 2011. - P. 1098-1106.
8. Meisner D., Gold B.T., Wenisch T.F. PowerNap: eliminating server idle power // CM SIGPLAN Notices. - 2009. - Vol. 44. - P. 205-216.
9. Nedeveschi S., Popa L., Iannaccone G., Ratnasamy S., Wetherall D. Reducing network energy consumption via sleeping and rate-adaptation // Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, NSDI'08, Berkeley, CA, USA. USENIX Association. - 2008. - P. 323-336.
10. Shorgin S.Y., Pechinkin A.V., Samouylov K.E., Gaidamaka Y.V., Gudkova I.A., Sopin E.S. Threshold-based Queuing System for Performance Analysis of Cloud Computing System with Dynamic Scaling // Proc. of the 12th International Conference of Numerical Analysis and Applied Mathematics ICNAAM-2014, Rhodes, Greece, 2014, USA, AIP Publishing - 2015. - Vol. 1648. - P. 1-3.

Поступила 21.10.2016

### Об авторах:

**Горбунова Анастасия Владимировна**, ассистент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов, agorbunova@sci.pfu.edu.ru;

**Самуйлов Константин Евгеньевич**, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов, доктор технических наук, профессор;

**Сопин Эдуард Сергеевич**, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управления» РАН, кандидат физико-математических наук.