

УДК 621.391.31

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ПРОПУСКНОЙ СПОСОБНОСТИ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ НА ЗАДЕРЖКУ ПРЕДОСТАВЛЕНИЯ СЕРВИСОВ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ SOA



Т.Н. КОВАЛЕНКО

Харьковский национальный
университет радиоэлектроники



С.А. МАКАРОВ

Харьковский университет воздушных
сил им. Ивана Кожедуба

Проведено аналіз ймовірнісних характеристик часу надання сервісів і вихідного потоку обслужених запитів в розподіленій системі із сервіс-орієнтованою архітектурою, а також залежності досліджених характеристик від пропускної здатності базової телекомунікаційної мережі в області низького та високого рівнів навантаження.

Analysis of probability characteristics of service provision time and output flow of served requests in a distributed system with service-oriented architecture is done, as well as dependence of these characteristics on throughput of basic telecommunication network at low and high load levels.

Проведен анализ вероятностных характеристик времени предоставления сервисов и выходного потока обслуженных запросов в распределенной системе с сервис-ориентированной архитектурой, а также зависимости данных характеристик от пропускной способности базовой телекоммуникационной сети в области низкого и высокого уровней нагрузки.

Введение

В последние годы одной из самых заметных тенденций в инфокоммуникационной отрасли стала реализация концепции «Все как услуга» (Everything as a service, EaaS), предполагающей предоставление абонентам инфраструктурных (Infrastructure as a service, IaaS), вычислительных и программных ресурсов (Software as a service, SaaS), средств хранения и обработки данных (Data as a service, DaaS) как услуг, доступных по сети Internet. Активный рост рынка подобных услуг наряду с ростом популярности GRID-технологий и технологий распределенных облачных вычислений выдвигает необходимость решения новых задач в области телекоммуникаций: очевидно, что телекоммуникационным сетям в обеспечении работоспособности подобных распределенных систем отводится одна из главных ролей. Вместе с тем, теоретическая, математическая база, позволяющая предложить эффективные, обоснованные решения по построению телекоммуникационной инфраструктуры таких систем значительно отстает от технологической.

Существующие на сегодняшний день модели распределенных систем, представленных, например, в работах [1-5], уделяют основное внимание корректности алгоритмов взаимодействия распределенных компонентов системы при реализации сложных последовательностей операций [3, 4], а телекоммуникационную систему

(ТКС), обеспечивающую данное взаимодействие, в лучшем случае представляют в виде канала передачи данных с постоянной задержкой [1, 5]. В связи с этим оценить производительность распределенной системы, качество реализуемых в ней сервисов и проанализировать их зависимость от характеристик реальной ТКС, на которой она строится, в настоящее время представляется возможным уже после реализации того или иного аппаратно-программного решения [6]. Большинство данных решений, предлагаемых такими известными производителями, как IBM [7], Hewlett-Packard [8], Sun Microsystems [9, 10], Microsoft [11], строятся на основе сервис-ориентированной архитектуры (Service Oriented Architecture, SOA). Результаты опроса, проведенного аналитической фирмой Forrester Research, показали, что 71% крупных компаний уже используют системы SOA или собираются внедрить их до конца 2011 года.

Архитектура SOA-систем и основные подходы к организации взаимодействия сервисов в них рассмотрены в работах [12-16]. Для обмена данными между сервисами, динамической маршрутизации запросов от прикладного компонента – потребителя сервиса и получения результатов от приложения – провайдера сервиса, а также для решения других коммуникационных задач в распределенных системах SOA используется технология корпоративной сервисной шины (Enterprise Service Bus, ESB) [12, 14]. ESB предоставляет единый механизм для передачи запросов и получения результатов сервисов, выполнения необходимых преобразований сообщений и транспортных протоколов, управления потоком обращений к сервисам и т.п. (рис. 1). Благодаря такому управлению выполняется необходимая последовательность вызовов сервиса для реализации бизнес-процесса.

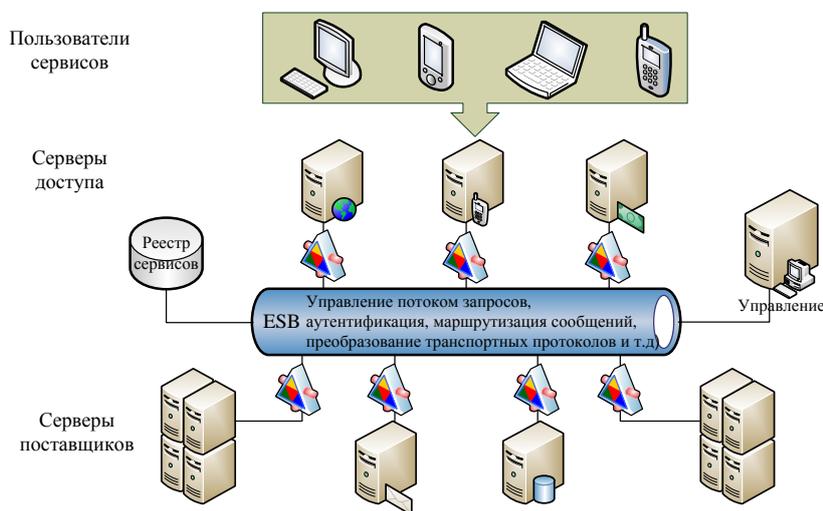


Рис. 1. Роль ESB в структуре SOA

Таким образом, характеристики базовой ТКС, обеспечивающей обмен сообщениями между территориально разнесенными серверами сервис-провайдеров и взаимодействие между компонентами распределенной SOA-системы, будут оказывать непосредственное влияние на характеристики системы в целом и качество предоставляемых ею сервисов [15, 16]. В связи с этим задача математического моделирова-

ния распределенных инфокоммуникационных систем SOA и исследования влияния пропускной способности базовой ТКС на задержку предоставления комплексного сервиса пользователю, решению которой посвящена данная работа, является в настоящее время актуальной.

I. Моделирование распределенных SOA-систем с использованием иерархических сетей Петри

Для количественной оценки влияния характеристик базовой ТКС на характеристики SOA-системы требуются адекватные математические модели, учитывающие телекоммуникационную составляющую таких систем. В данной работе для математического описания распределенной системы SOA были применены раскрашенные временные иерархические сети Петри (СП). СП представляют собой математический аппарат моделирования и анализа стохастических динамических систем и процессов [17 - 19]. Иерархическая раскрашенная СП (CPN) представляет собой совокупность нескольких CPN модулей, объединенных в единую модель сложной системы с помощью специальных переходов и позиций. Существующие модули могут использоваться в модели несколько раз, кроме того, на их основе можно создавать новые модули. Иерархическая структура такой модели может быть сколь угодно сложной – каждый модуль может состоять из нескольких более мелких CPN модулей, модули разного уровня иерархии могут иметь общие элементы (позиции слияния).

При построении модели сетевого взаимодействия распределенных компонентов SOA-системы был применен подход «сверху вниз», который предполагает построение обобщенной модели системы с постепенной детализацией моделей процессов, требующих более тщательного анализа. В этом случае любой переход при необходимости может быть преобразован в замещающий, а детальное описание соответствующего данному переходу процесса представляется отдельным CPN модулем [19]. Иерархическая структура предлагаемой модели приведена на рис. 2.

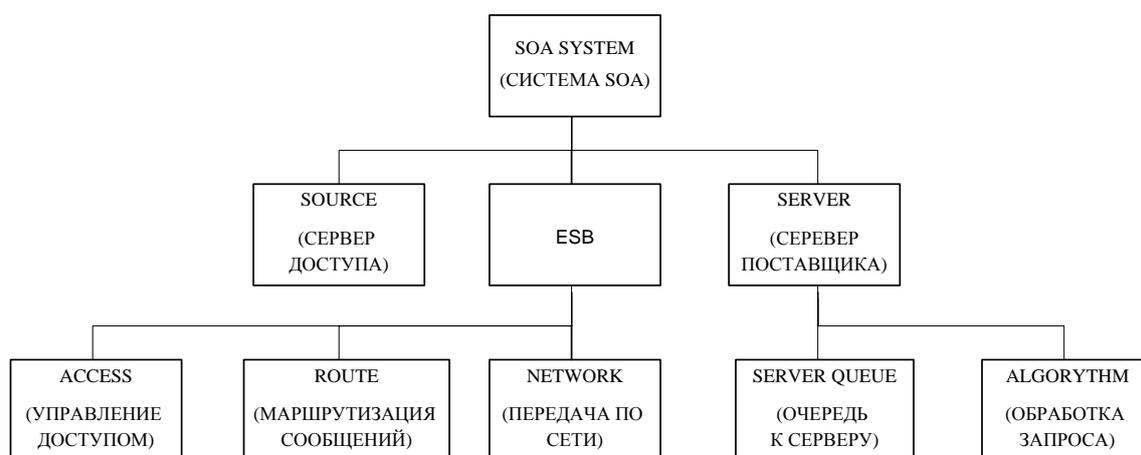


Рис. 2. Иерархическая структура модели взаимодействия компонентов SOA-системы

Верхний уровень модели представляет собой CPN модуль, описывающий наиболее общие аспекты построения системы SOA, ее структуру и функциональные

компоненты. Обобщенная модель SOA-системы (SOA System) представлена на рис. 3, а. Основными компонентами модели являются замещающие переходы «Source» (сервер доступа пользователей к сервисам, предоставляемых системой), «ESB» и «Server» (сервер поставщика сервисов), каждый из которых представляет собой CPN модуль следующего более низкого уровня иерархии. Данные CPN модули были описаны ранее в работе [20].

CPN модуль сервера доступа к сервисам (SOURCE) описывает процесс формирования запросов к системе SOA. В данном модуле происходит моделирование входного трафика, задаются размер и интенсивность передаваемых по сети блоков данных. Сбор статистической информации для ее дальнейшей обработки и анализа характеристик моделируемой системы осуществляется с помощью позиций Amount и ServiceTime, на которых размещаются фишки успешно обработанных запросов на сервисы и времени обработки данных запросов соответственно.

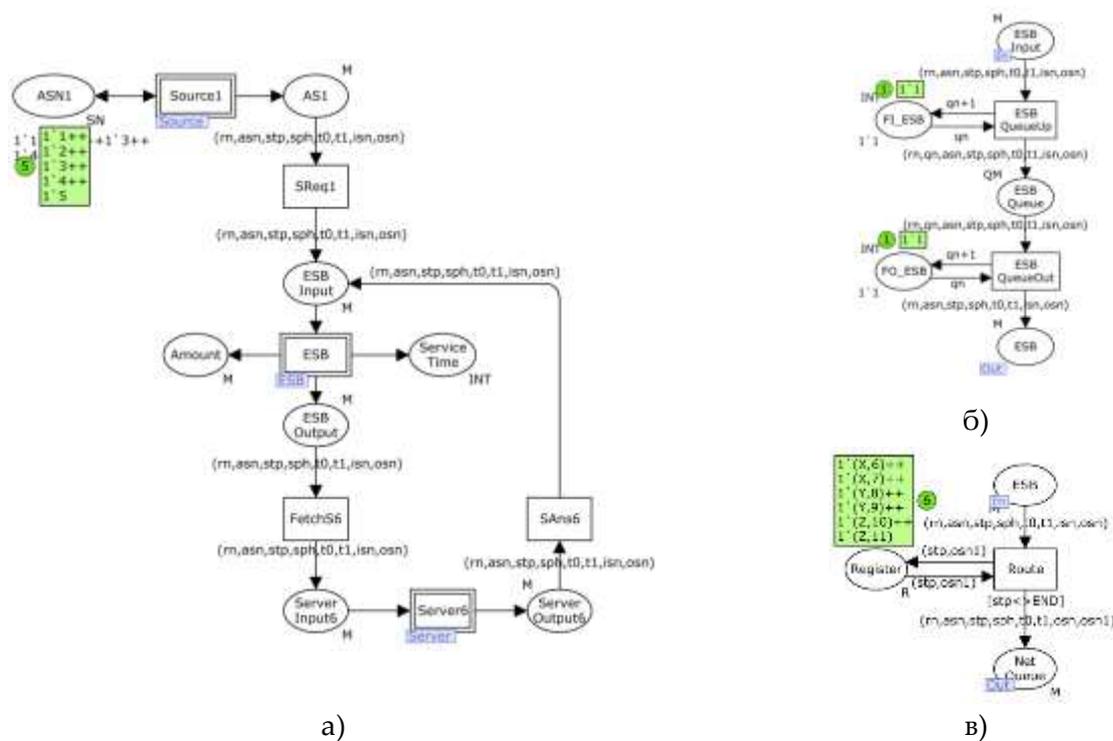


Рис. 3. Обобщенная модель SOA-системы (а), CPN модуль управления доступом (б) и маршрутизации сообщений (в)

CPN модуль корпоративной сервисной шины (ESB) состоит из трех основных модулей, реализующих функции ESB: управление входным потоком запросов к сервисной шине (ограничение потока, порядок обслуживания очереди запросов и т.д.) моделирует CPN модуль ACCESS (рис. 3, б); процедуры маршрутизации принятого к обслуживанию запроса к серверу поставщика моделирует CPN модуль ROUTE (рис. 3, в), который может реализовывать, например, процедуру оптимального выбора поставщика сервиса и распределения сетевых ресурсов при обслуживании потоков запросов на сервисы [21]; процесс передачи сообщений-запросов и ответов на них по сети и возникающие при этом задержки моделирует CPN модуль Network (рис. 4).

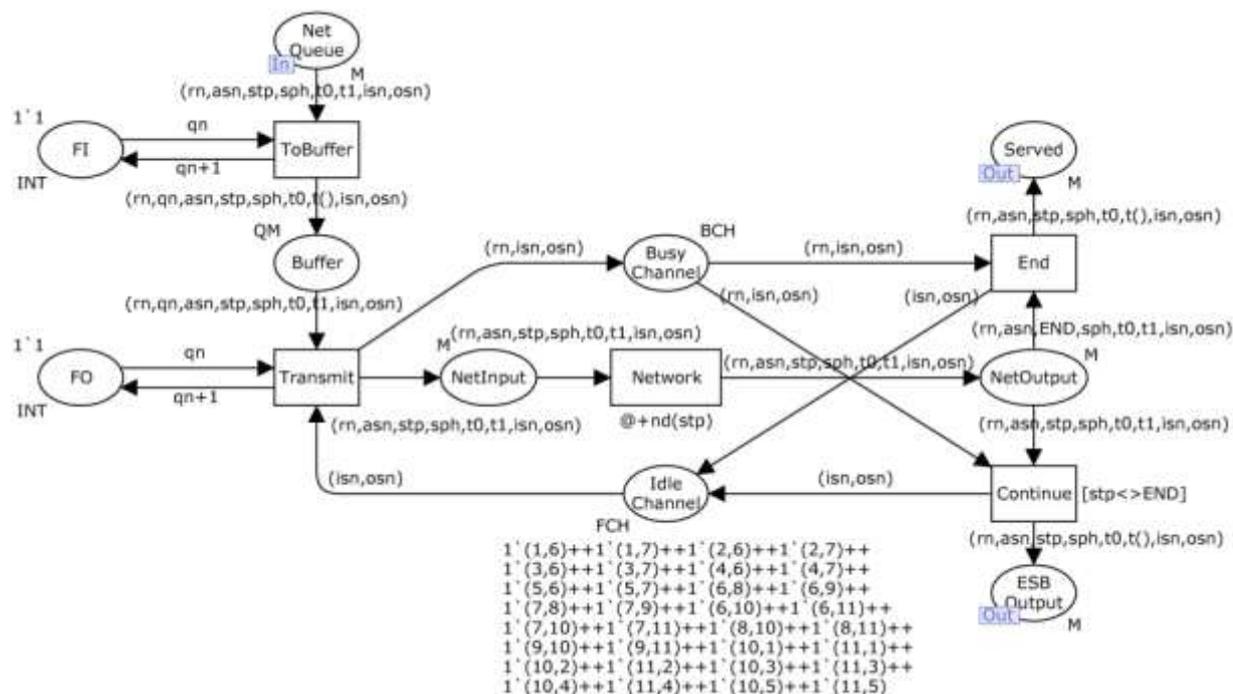


Рис. 4. Модель передачи сообщений по сети

Модель сервера поставщика сервисов представлена в работе [3], она также включает два CPN модуля, описывающих процедуры обслуживания входной очереди запросов на каждом сервере (модуль Server Queue) и процесс обработки запросов на сервисы различного типа с учетом возникающих при этом задержек (модуль Algorithm). Представленная модель процессов сетевого взаимодействия в распределенных системах SOA позволяет проводить анализ таких характеристик, как время обслуживания запросов, размеры очередей, производительность моделируемой системы, вероятностные характеристики выходных потоков обслуженных запросов и др.

II. Анализ задержек предоставлении сервисов в распределенной SOA-системе

С целью исследования влияния пропускной способности базовой ТКС на задержку предоставления сервисов в распределенных SOA-системах была проведена серия экспериментов по имитационному моделированию с применением предложенной модели системы SOA и программного пакета моделирования сетей Петри «CPN Tools». В ходе исследования были проанализированы такие характеристики, как полное время предоставления сервиса, производительность системы SOA, выходной поток обслуженных запросов. При этом рассматривалось функционирование системы под воздействием трех типов входного потока запросов на сервисы и при различном уровне нагрузки (коэффициент загрузки системы варьировался в пределах $0,1 < \rho < 1$). Характеристики входных потоков и другие исходные данные проведенного моделирования приведены в табл. 1.

Таблица 1. Исходные данные

Среднее значение интенсивности входного потока запросов на сервисы от одного сервера доступа, λ , запр./с		1...10
Интервал времени между запросами во входном потоке	Тип потока 1	фиксированный
	Тип потока 2	случайный, экспоненциально распределенный
	Тип потока 3	случайный, с равномерным законом распределения
Пропускная способность телекоммуникационной сети между серверами сервис-провайдеров, C , Мбит/с		10...50
Закон распределения времени передачи сообщений по сети		экспоненциальный
Размер сообщений, кбайт		125
Производительность серверов сервис-провайдеров, m , запр./с		20...100
Закон распределения времени обработки запроса на сервере сервис-провайдера		экспоненциальный

Топология моделируемой распределенной системы взята аналогичная рассмотренной в работе [21]: в состав системы входят пять серверов доступа к сервисам, шесть серверов сервис-провайдеров, предлагающих три различных набора услуг (по два сервера с аналогичным набором сервисов). Интенсивности входных потоков запросов на сервисы на всех серверах доступа, производительности серверов и значения пропускной способности ТКС между ними принимались одинаковыми.

Анализ результатов моделирования показал, что независимо от типа рассмотренных входных потоков плотность распределения вероятностей задержек предоставления сервиса системой SOA может быть представлена гамма-распределением (проверка правдоподобия гипотез о законах распределения исследованных величин была проведена с использованием критерия χ^2 Пирсона):

$$f(t) = \frac{\left(\frac{t}{a}\right)^{b-1}}{a \cdot \Gamma(b)} \cdot \exp\left(-\frac{t}{a}\right),$$

где a – масштабный коэффициент; b – параметр формы; $\Gamma(b)$ – гамма-функция.

Значения параметров a и b данного распределения при одинаковых исходных данных модели для различных типов входных потоков отличались, однако для всех трех типов исследованных потоков было установлено, что при увеличении коэффициента загрузки системы значение масштабного коэффициента a увеличивается, а значение параметра формы b уменьшается. Результаты анализа вероятностных характеристик задержек предоставления сервисов в распределенной SOA-системе сведены в табл. 2, а на рис. 5 представлены типичные гистограммы плотности распределения вероятностей задержки предоставления сервисов в области низкого ($\rho < 0,2$, рис. 5, а) и высокого ($\rho > 0,7$, рис. 5, б) уровней нагрузки на систему. Данные гистограммы получены для входного потока типа 3 при $C = 50$ Мбит/с, $m = 100$ запр./с, $\lambda = 1$ запр./с и $\lambda = 7$ запр./с соответственно.

На рис. 6 приведены графики зависимости среднего значения задержки предоставления сервисов в распределенной системе SOA от пропускной способности базовой ТКС (рис. 6, а) и интенсивности входного потока запросов (рис. 6, б) для трех исследованных в работе типов потоков.

Таблица 2. Вероятностные характеристики задержки предоставления сервисов

Характеристики	Тип потока 1	Тип потока 2	Тип потока 3
Плотность распределения вероятности	гамма-распределение		
Масштабный коэффициент a	увеличивается с ростом нагрузки		
Параметр формы b	уменьшается с ростом нагрузки		
Значение b в области низких уровней нагрузки ($\rho < 0,2$)	$b > 7$	$b > 6$	$b > 6$
Значение b в области высоких уровней нагрузки ($\rho > 0,7$)	$b = 3,5...4,5$	$b = 2...4$	$b = 2...4$

Из полученных результатов можно сделать вывод, что увеличение пропускной способности базовой ТКС между серверами сервис-провайдеров позволяет сократить общее время обслуживания запроса в распределенной SOA-системе, однако степень зависимости задержки предоставления сервиса от пропускной способности ТКС для различных типов входных потоков несколько отличается: так, наибольший выигрыш был получен для потока типа 2, в то время как для потока типа 1 сокращение задержки предоставления сервиса оказалось наименьшим.

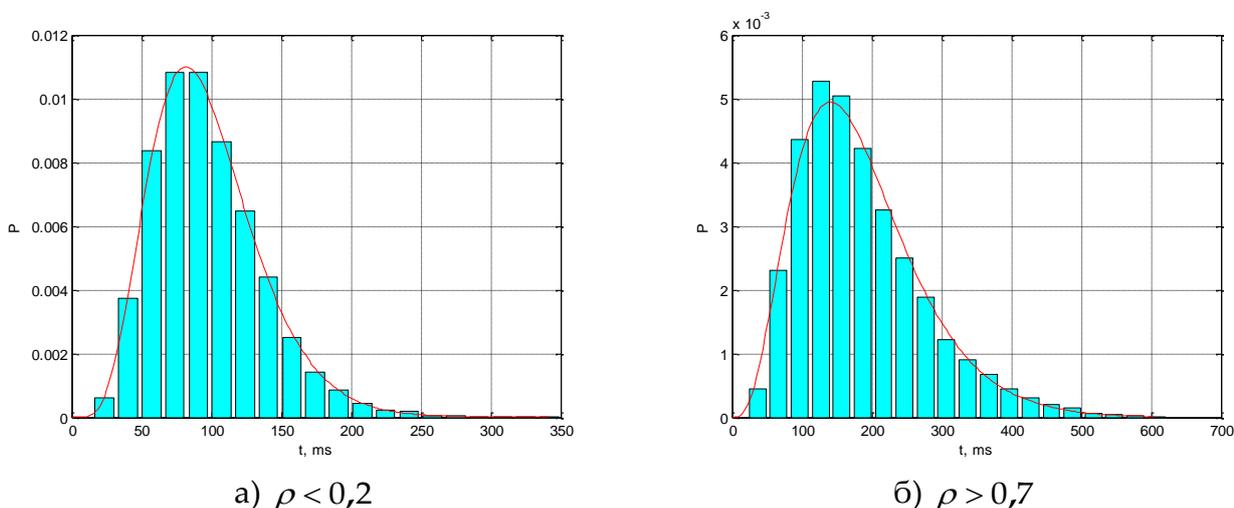


Рис. 5. Гистограммы плотности распределения вероятностей задержки предоставления сервисов

Необходимо также отметить, что для заданного значения производительности серверов сервис-провайдеров ($m = 100$ запр./с) увеличение пропускной способности ТКС выше 30 Мбит/с нерационально, т.к. это не приведет к ощутимому снижению времени обслуживания запросов в связи с тем, что «узким местом» в рассмотренной SOA-системе будет являться уже не сеть, а серверы, обрабатывающие запросы.

Графики, приведенные на рис. 6, б, иллюстрируют, что при невысоком уровне нагрузки на исследуемую систему (значение интенсивности входного потока запросов от одного сервера доступа $\lambda = 5$ запр./с при пропускной способности базовой ТКС $C = 50$ Мбит/с и производительности серверов сервис-провайдеров $m = 100$ запр./с соответствует коэффициенту загрузки системы $\rho \approx 0,5$) наибольшее значение средней задержки предоставления сервиса наблюдается в случае входного потока типа 1. В то же время при увеличении интенсивности входного потока средняя задержка предоставления сервиса наиболее резко возрастает для входного потока типа 2, и в области среднего и высокого уровней нагрузки ($\rho > 0,5$) средняя задержка предоставления сервиса в распределенной SOA-системе является наибольшей именно для этого типа входного потока.

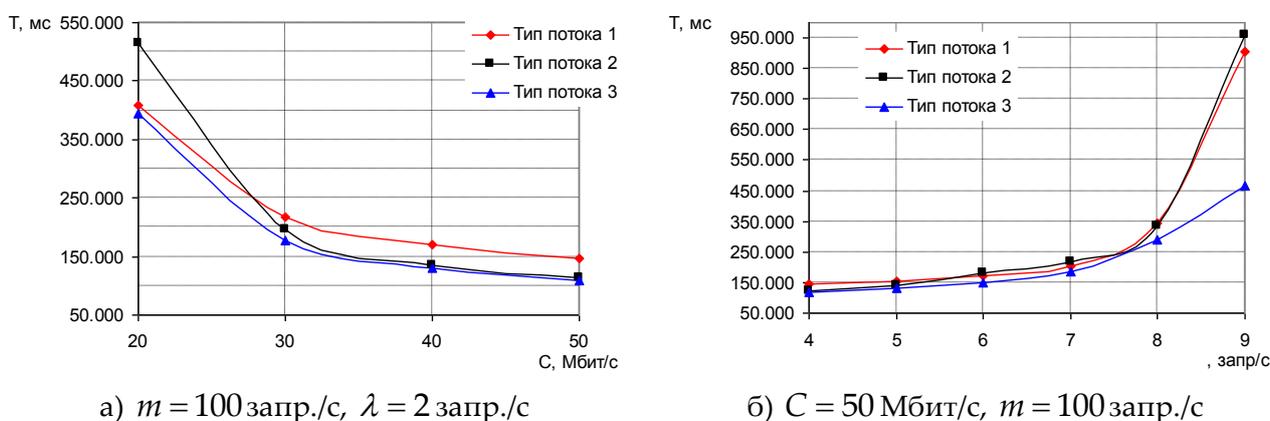


Рис. 6. Средняя задержка предоставления сервисов в распределенной системе SOA

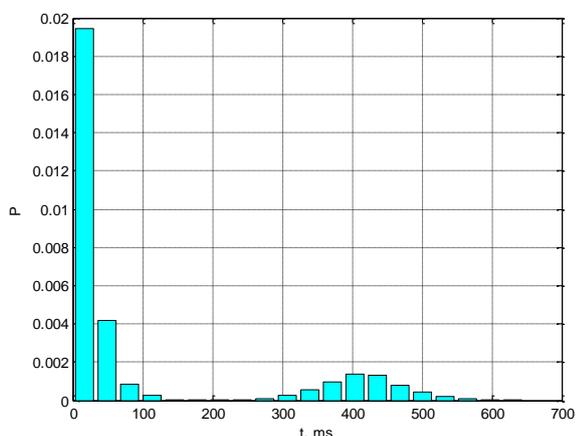
Отметим, что при проведении моделирования все входные потоки не полагались ординарными. В частности, в случае детерминированного входного потока типа 1 вследствие того, что интервал времени между поступлением запросов на сервисы полагался фиксированным и одинаковым для всех серверов доступа, запросы от нескольких серверов поступали в систему одновременно, «пачками», чем и объясняются достаточно высокие значения суммарного времени обслуживания запросов системой по сравнению со случайными входными потоками типа 2 и 3.

В ходе исследований были также проанализированы характеристики выходных потоков обслуженных запросов в распределенной SOA-системе. Результаты проведенного анализа сведены в табл. 3, а на рис. 7 приведены типичные гистограммы плотности распределения вероятностей интервалов времени между обслуженными запросами в выходном потоке в случае входного потока типа 1, полученные для значений $C = 50$ Мбит/с, $m = 100$ запр./с, $\lambda = 2$ запр./с и $\lambda = 6$ запр./с соответственно. Таким образом, было установлено, что в случае входных потоков типа 2 и 3 интервалы времени между обслуженными запросами в выходном потоке имеют гамма-распределение, причем для потока типа 2 в области невысокого уровня нагрузки ($\rho \leq 0,3$) параметр формы гамма-распределения $b \approx 1$, т.е. распределение

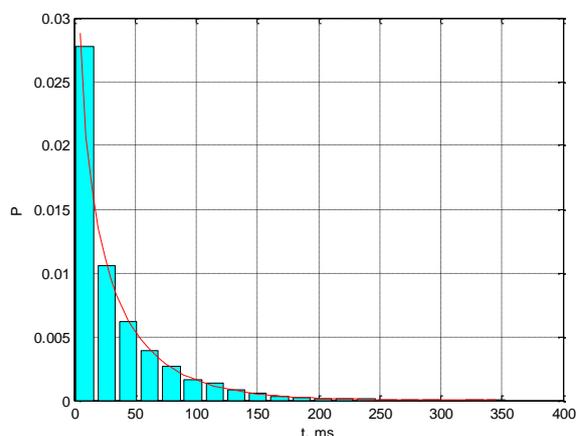
близко к экспоненциальному, а с ростом интенсивности входного потока оба параметра a и b гамма-распределения уменьшаются.

Таблица 3. Характеристики выходного потока обслуженных запросов

Характеристики		Тип потока 1	Тип потока 2	Тип потока 3
Закон распределения интервалов времени между обслуженными запросами	$\rho < 0,2$	плотность распределения вероятностей имеет сложную бимодальную форму	экспоненциальный	гамма-распределение
	$\rho > 0,7$	гамма-распределение	гамма-распределение	гамма-распределение
Параметр формы гамма-распределения b		0,5...0,7	0,4...1,0	0,5...1,2
уменьшается с ростом нагрузки				



а) $\lambda = 2$ запр./с, $\rho \leq 0,3$



б) $\lambda = 6$ запр./с, $\rho \approx 0,7$

Рис. 7. Гистограммы плотности распределения вероятностей интервалов времени между обслуженными запросами (входной поток типа 1)

Что касается детерминированного входного потока типа 1, при невысоком уровне нагрузки ($\rho \leq 0,3$) плотность распределения вероятностей интервалов времени между обслуженными запросами в выходном потоке имеет сложную бимодальную форму (рис. 7, а), а с ростом нагрузки распределение становится близким к гамма-распределению, хотя мера расхождения гипотетического и экспериментально полученного распределения несколько выше, чем для входных потоков типа 2 и 3.

Выводы

В работе представлены результаты исследования задержек предоставления сервисов в распределенных системах с сервис-ориентированной архитектурой. Проведен анализ вероятностных характеристик задержки предоставления сервисов и их зависимости от пропускной способности базовой ТКС, обеспечивающей обмен сообщениями между распределенными компонентами SOA-системы (серверами сер-

вис-провайдеров). Были также исследованы характеристики выходных потоков обслуженных запросов. Полученные результаты позволяют сформулировать следующие основные выводы о функционировании распределенной системы SOA под воздействием рассмотренных детерминированного и недетерминированных потоков запросов при различных уровнях нагрузки (коэффициент загрузки $0,1 < \rho < 1$):

1. Суммарная задержка предоставления сервиса в исследованной распределенной системе SOA имеет гамма-распределение, при увеличении коэффициента загрузки системы (уменьшении пропускной способности базовой ТКС) значение масштабного коэффициента гамма-распределения a увеличивается, а значение параметра формы b уменьшается.

2. Наиболее существенное сокращение задержек предоставления сервиса в распределенной SOA-системе за счет увеличения пропускной способности базовой ТКС (при прочих равных условиях) наблюдается для случайного входного потока с экспоненциально распределенными интервалами времени между запросами, а наименьший выигрыш наблюдается в случае детерминированного неординарного входного потока запросов.

3. При невысоком уровне нагрузки (коэффициент загрузки системы $\rho < 0,5$) наибольшее значение средней задержки предоставления сервиса в распределенной SOA-системе наблюдается в случае детерминированного неординарного входного потока, а при $\rho > 0,5$ средняя задержка предоставления сервисов является наибольшей для случайного входного потока с экспоненциально распределенными интервалами времени между запросами.

4. При высоком уровне нагрузки на систему ($\rho > 0,7$) плотность распределения вероятностей интервалов времени между обслуженными запросами в выходном потоке имеет гамма-распределение для всех трех исследованных типов входных потоков запросов. С увеличением интенсивности входного потока параметр формы гамма-распределения b уменьшается.

5. При коэффициенте загрузки системы $\rho \leq 0,3$ плотность распределения вероятностей интервалов времени между обслуженными запросами в выходном потоке имеет сложную бимодальную форму (в случае детерминированного входного потока), подчиняется гамма-распределению (в случае равномерно распределенного интервала времени между поступающими в систему запросами) или близка к экспоненциальному распределению (если интервалы времени между запросами во входном потоке являются экспоненциально распределенными).

Направлением дальнейших исследований является анализ производительности распределенных систем SOA под воздействием самоподобного входного потока запросов, исследование влияния потерь и отказов в базовой телекоммуникационной сети на характеристики распределенных SOA-систем, а также разработка и анализ эффективности методов управления трафиком в распределенных инфокоммуникационных системах.

Список литературы:

1. *Топорков В.В.* Модели распределенных вычислений. — М.: Физматлит, 2004. — 320 с.
2. *Lee E.A., Sangiovanni-Vincentelli A.* A framework for comparing models of computations // IEEE Trans. On CAD of Integrated Circuits and Systems. — 1998. — Vol.17, No.12. — P. 1217 – 1229.
3. *Воеводин В.В., Воеводин Вл.В.* Параллельные вычисления. — СПб.: БХВ-Петербург, 2002. — 608 с.
4. *Корнеев В.В.* Параллельные вычислительные системы. — М.: Нолидж, 1999. — 320 с.
5. *Banner Th., Ernst R., Osterlinng A.* Scalable performance scheduling for hardware-software cosynthesis // European Design Automation Conf.: Proc. — Brighton, UK: IEEE CS Press, 1995. — P. 164 – 169.
6. *Хойссер М.* Тестирование SOA: примеры [Электронный ресурс] // Директор информационной службы. — 2008. — № 12. — Режим доступа: <http://www.osp.ru/cio/2008/12/5585238/>.
7. Working smarter with a Smart SOA foundation [Электронный ресурс]. — Режим доступа: ftp://public.dhe.ibm.com/software/solutions/pdfs/7946_Working_smarter_Brochure.pdf.
8. HP SOA Systinet and JBoss Enterprise SOA Platform. Service-oriented architecture infrastructure and governance solution [Электронный ресурс]. — Режим доступа: <http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA2-6915ENW.pdf>.
9. *Ed Ort.* Service-Oriented Architecture and Web Services: Concepts, Technologies, and Tools [Электронный ресурс]. — Режим доступа: <http://java.sun.com/developer/technicalArticles/WebServices/soa2/SOATerms.html>.
10. Oracle White Paper — Oracle SOA Suite 11g [Электронный ресурс]. — Режим доступа: <http://www.oracle.com/us/technologies/soa/soa-suite/oracle-soa-suite-11g-198758.pdf>.
11. Microsoft and SOA Software [Электронный ресурс]. — Режим доступа: http://www.soa.com/solutions/microsoft/microsoft_and_soa_software/.
12. *Пузыня С., Пырлина И.* Корпоративная архитектура и SOA [Электронный ресурс] // Открытые системы. — 2010. — №2. — Режим доступа: <http://www.osp.ru/os/2010/02/13001450/>.
13. *Nicolai M. Josuttis.* SOA in Practice: The Art of Distributed System Design (Theory in Practice). — O'Reilly Media, 2007. — 352 p.
14. *Дубова Н.* SOA: подходы к реализации [Электронный ресурс] // Открытые системы. — 2004. — № 6. — Режим доступа: <http://www.osp.ru/os/2004/06/184450/>.
15. *Сергей Кузнецов.* SOA с гарантией качества [Электронный ресурс] // Открытые системы. — 2008 — № 7. — Режим доступа: <http://www.osp.ru/os/2008/07/5479587/>.
16. *Брент Слипер.* 5 проблем SOA [Электронный ресурс] // Computerworld Россия. — 2004. — № 38. — Режим доступа: <http://www.osp.ru/cw/2004/38/81416/>.
17. *Popovskij V., Barkalov A., Titarenko L.* Control and Adaptation in Telecommunication Systems: Mathematical Foundations. — Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. — 564 p.
18. Математичні основи теорії телекомунікаційних систем / В.В. Поповський, С.О. Сабурова, В.Ф. Олійник, Ю.І. Лосев, Д.В. Агеев та ін. За загальною редакцією В.В. Поповського. — Харків: ТОВ «Компанія СМІТ», 2006. — 564 с.
19. *Jensen K.* Coloured Petri nets: basic concepts, analysis methods and practical use. — Berlin: Heidelberg, New York: Springer-Verlag, 1996. — 234 p.
20. *Коваленко Т.Н.* Методика моделирования процессов сетевого взаимодействия в системах с сервис-ориентированной архитектурой // Радиотехника. — Вып.163. — С. 66 – 72.
21. *Kovalenko T.* Optimization of network resources distribution in SOA systems // Journal of EcoAgriTourism. — 2010. — Vol. 6, № 2(19). — P. 164 – 167.