

Общероссийский математический портал

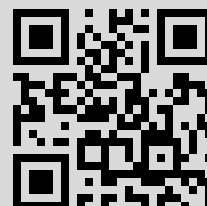
Д. В. Жевнерчук, А. В. Николаев, Методика моделирования нагрузки на сервер в открытых системах облачных вычислений, *Информ. и её примен.*, 2012, том 6, выпуск 2, 43–50

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 217.118.78.110

30 апреля 2016 г., 20:07:25



МЕТОДИКА МОДЕЛИРОВАНИЯ НАГРУЗКИ НА СЕРВЕР В ОТКРЫТЫХ СИСТЕМАХ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ

Д. В. Жевнерчук¹, А. В. Николаев²

Аннотация: Планирование серверного ресурса систем облачных вычислений является сложной задачей. Необходимо учитывать такие факторы, как состав и параметры аппаратной платформы, параметры системного программного обеспечения, управляющего выполнением прикладных программ, свойства трафика, порождаемого пользователями и определяющего режимы функционирования прикладных программ. Предложена методика оценки загрузки серверного ресурса открытых систем облачных вычислений (ОСОВ) на основе анализа процессов взаимодействия пользователей с программным обеспечением. Обоснована ее достоверность. Приведены результаты моделирования нагрузки на серверную часть системы управления средами имитационного моделирования.

Ключевые слова: облачные вычисления; имитационное моделирование; человеко-машинное взаимодействие

1 Введение

Вопросы проектирования систем поддержки удаленного вычислительного эксперимента до конца не решены. В ходе проектирования ОСОВ возникает задача моделирования потока запросов, приводящих к загрузке серверной части. Такие модели позволяют получить оценку аппаратного ресурса для обслуживания некоторого количества клиентских систем при рабочей и пиковой нагрузке. В работах [1–3] рассматриваются теоретические модели трафика в локальных и глобальных сетях. В основном полученные результаты имеют практическую значимость при проектировании средств передачи данных в сети. В работах [3–6] построены модели трафика, поступающего на вход серверов разного типа, таких как веб-серверы, серверы баз данных и др. На основании обзора работ были сделаны следующие выводы:

1. Модели описывают трафик систем, построенных на основе определенных технологий и/или предназначенных для решения ограниченного круга задач.
2. Процессы формирования запросов моделируются на основании замеров уже переданного в сеть трафика.
3. Модели описывают смешанный трафик.

Для анализа трафика ОСОВ классические методики моделирования трафика не эффективны, поскольку

- в общем случае ОСОВ обладает свойствами расширения по произвольным программно-аппаратным платформам, по решаемым задачам, по источникам нагрузки;
- в ОСОВ постоянно происходят качественные изменения, поэтому на основании конечного числа измерений трафика можно построить модели, описывающие ОСОВ только в некотором подмножестве состояний;
- отсутствуют развитые средства автоматизации и механизмы контроля перехода ОСОВ в новое качественное состояние.

Таким образом, задача разработки эффективных методик оценки нагрузки ОСОВ до конца не решена и является актуальной.

2 Постановка задачи

Была поставлена задача разработки методики моделирования и построения на ее основе моделей нагрузки на серверную часть в ОСОВ. К методике и моделям предъявлены следующие требования:

1. Модели должны описывать дифференцированный трафик, из которого можно выделить потоки, принадлежащие определенному программному обеспечению и связанные с определенными задачами.
2. Источник дифференцированного трафика должен определяться процессами человеко-машинного взаимодействия.

¹Чайковский технологический институт (филиал) Ижевского государственного технического университета, drevnigeck@yandex.ru

²Чайковский технологический институт (филиал) Ижевского государственного технического университета, elodssa@yandex.ru

3. В модель должны передаваться эмпирические функции распределения вероятностей интервалов времени между передачами управляющих сигналов серверу, приводящих к существенной загрузке центрального процессора и оперативной памяти.
4. Должен проводиться системный анализ процессов решения пользовательских задач с применением программного обеспечения и учетом поведения пользователя при решении задач с помощью программ.
5. Должна обеспечиваться высокая степень автоматизации процессов сбора эмпирических данных.

Ставилась задача применить представленную методику для изучения системы обработки сред имитационного моделирования и проверить достоверность построенных моделей нагрузки на сервер.

3 Методика моделирования нагрузки на сервер

Предлагаемая методика моделирования нагрузки на сервер включает ряд этапов:

1. Сбор сведений о процессе взаимодействия клиента и сервера.
2. Определение последовательности выполнения действий.
3. Построение имитационной модели процесса взаимодействия.
4. Проведение экспериментов с имитационной моделью процесса взаимодействия и необходимой настройки.
5. Адаптацию полученного генератора нагрузки к работе с внешней системой.

Схема методики представлена на рис. 1.

На первом этапе необходимо собрать сведения о взаимодействии клиента и сервера. Клиент работает с сервером в режиме запрос–ответ. Необходимо получить данные об интервалах времени между определенными действиями клиента. Для упрощения сбора данных было разработано клиент-серверное приложение «Хронометр», позволяющее настроить список действий пользователя, требующих отметки времени выполнения. После настройки клиентская часть «Хронометра» начинает замерять интервалы времени, в течение которых выполняются действия, и передавать собираемые данные на сервер. Такой подход упрощает сбор

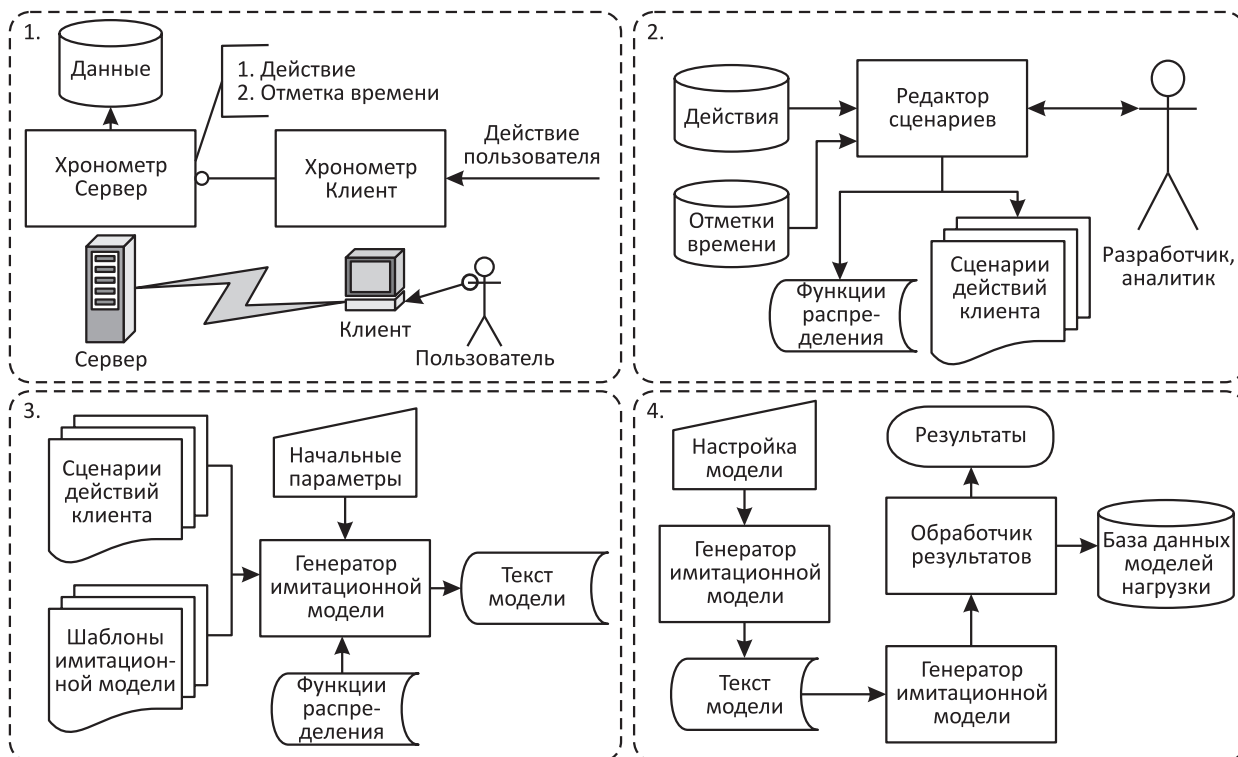


Рис. 1 Методика моделирования нагрузки на сервер

сведений, так как данные можно собирать сразу с группы пользователей.

На втором этапе определяется последовательность действий клиента по отношению к серверу. Для этого выполняются следующие шаги. На основе собранных данных строятся эмпирические функции распределения вероятностей интервалов времени, затраченного на действия клиента. Строятся сценарии человеко-машинного взаимодействия и вводится классификация пользователей, использующих определенные сценарии. Для упрощения работы было разработано программное средство «Редактор сценариев», позволяющее быстро обработать список действий пользователя и увидеть максимальное, минимальное и среднее время выполнения действий. Кроме того, с его помощью можно построить эмпирические функции распределения интервалов времени в синтаксисе языка GPSS (General Purpose Simulation System).

На третьем этапе строится имитационная модель процесса взаимодействия клиентов с сервером для оценки времени между событиями прихода запросов от клиента, приводящих к существенной загрузке центрального процессора и оперативной памяти. Для автоматизации построения имитационной модели было разработано программное средство «Генератор имитационной модели» (ГИМ). С его помощью можно быстро получить код модели на основании вводимых параметров, эмпирических функций и сценариев действий. Программное средство ГИМ использует заготовленные шаблоны на языке имитационного моделирования GPSS.

Далее проводится эксперимент с имитационной моделью, определение необходимых параметров и установка настроек. После этого строится модель трафика, поступающего на сервер, включающего запросы от клиента, приводящие к существенной загрузке центрального процессора и оперативной памяти. Все настройки и текст модели сохраняются

в базу данных для последующего воспроизведения потока запросов.

На последнем этапе полученную модель генерации запросов пользователя адаптируют для работы с внешней системой.

4 Ход исследования

Исследования были проведены для среды моделирования GPSS World Student, с которой пользователи работают в режиме обучения. В построенных моделях трафика учитывалась информация о запросах, приводящих к прогонам имитационных моделей, что влияет на загрузку аппаратного ресурса. Была проверена гипотеза о достоверности построенных моделей.

4.1 Сбор данных о процессе взаимодействия учащегося со средой GPSS World Student

Был проведен хронометраж действий учащихся по изучению среды моделирования с помощью учебных моделей, по кодированию и отладке модели. Для формирования журнала действий пользователя использовалось программное средство «Хронометр 1.0».

Было замечено, что в ходе выполнения учебного задания учащийся сначала формирует код модели, далее модель тестируется и отлаживается. Это сопровождается определенным числом попыток компиляции модели и интервалами времени поиска и устранения ошибок в коде. После отладки выполняется разовый контрольный прогон модели, после чего анализируется выходной отчет, на основании которого осуществляется поиск логических ошибок. Учащимся может быть проведено несколько дополнительных исправлений кода. Модель вновь

Таблица 1 Наблюдение за процессом изучения GPSS World Student

Действие	Тип моделей	Категория	
		Успевающие	Неуспевающие
Количество ошибок компиляции в режиме отладки (1 задание)	Простые модели	[0–3]	[2–6]
	Сложные модели	[4–12]	[8–16]
Поиск ошибки и ее устранение, с		[30–120]	[90–200]
Анализ итогового отчета, с	Простые модели	[20–60]	[40–120]
	Сложные модели	[20–120]	[90–240]
	Первичное ознакомление	[60–120]	[90–120]
Кодирование новой модели, мин		[120–300]	[240–420]
(подготовка первого варианта кода модели), мин		[360–720]	[600–1080]
Работа со средой моделирование по инструкции (время поиска функциональности)		[20–40]	[30–90]

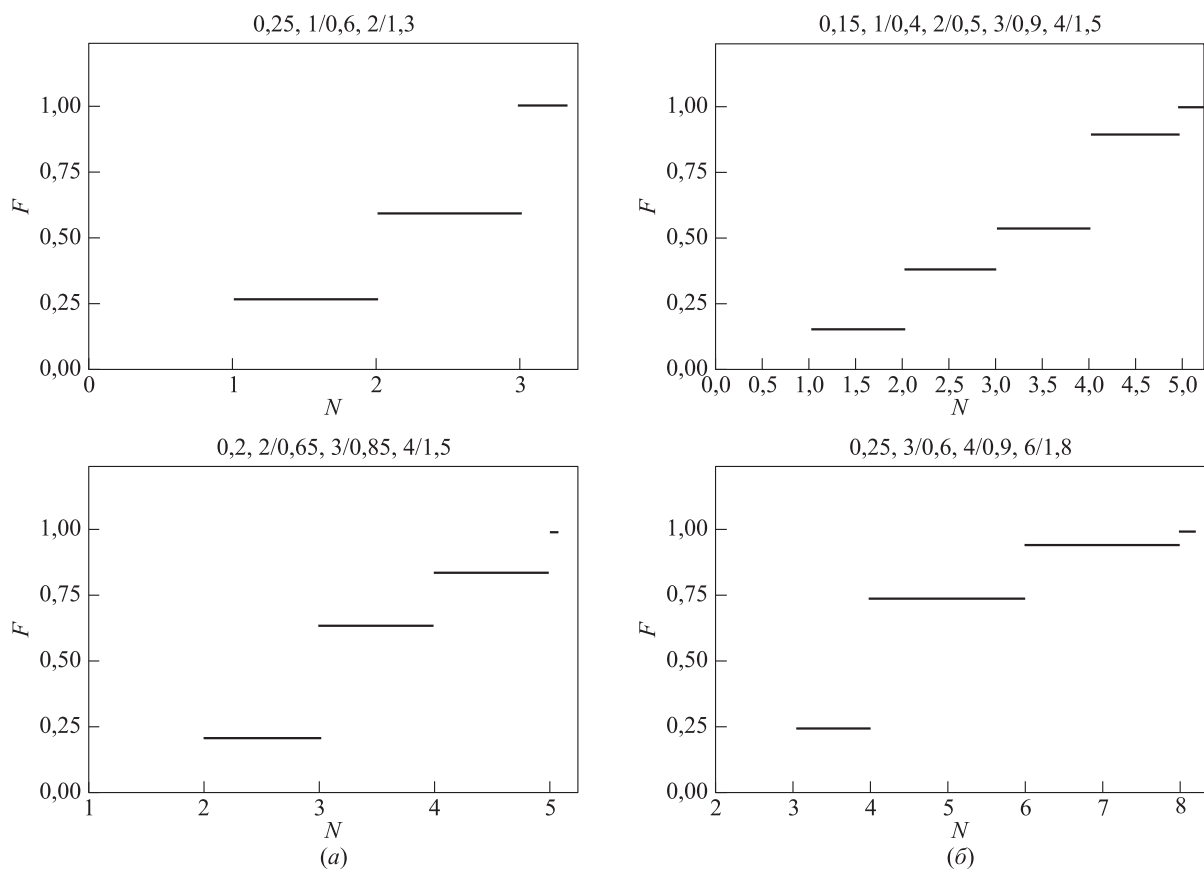


Рис. 2 Число шагов отладки простой (а) и сложной (б) модели успевающим (верхний ряд) и неуспевающим учащимся (нижний ряд)

тестируется, отлаживается, и результаты итогового прогона снова анализируются.

В режиме изучения готовой модели или среды моделирования с использованием методических указаний основное время тратится на анализ инструкций.

Наблюдения проводились за тремя учебными группами общей численностью 43 чел. Было проведено 5 занятий (10 академических часов). На основе полученных данных построена классификация учащихся и задач по времени решения. Все учащиеся разделены на две группы: «успевающие» и «неуспевающие», а задачи — на группы «простые» и «сложные». Полученные граничные оценки интервалов времени действий учащегося в режимах обучения и выполнения задания приведены в табл. 1.

С помощью программы «Редактор сценариев» были построены необходимые эмпирические законы распределения интервалов времени (рис. 2—5).

Полученные результаты были использованы при построении имитационных моделей взаимодействия учащихся со средой GPSS World.

4.2 Модель взаимодействия учащегося со средой GPSS World Student

Выделим события, приводящие к компиляции и прогону модели, а следовательно, и к загрузке центрального процессора. При возникновении события первого запуска модели (e_1) происходит компиляция, в результате которой формируется отчет о готовности прогона. При возникновении события «запуск модели» (e_2) выполняется прогон, в результате которого формируется отчет с откликом. В модели вводится 2 класса задач: простые и сложные, — и 2 класса учащихся: успевающие и неуспевающие. В зависимости от класса задачи и типа учащегося выбираются построенные в результате хронометража функции, определяющие количество ошибок компиляции и время задержки при написании кода модели, анализа ошибок компиляции, анализа итогового отчета. При построении модели были сделаны следующие допущения:

1. Время компиляции модели пренебрежимо мало по сравнению со временем прогона модели, а

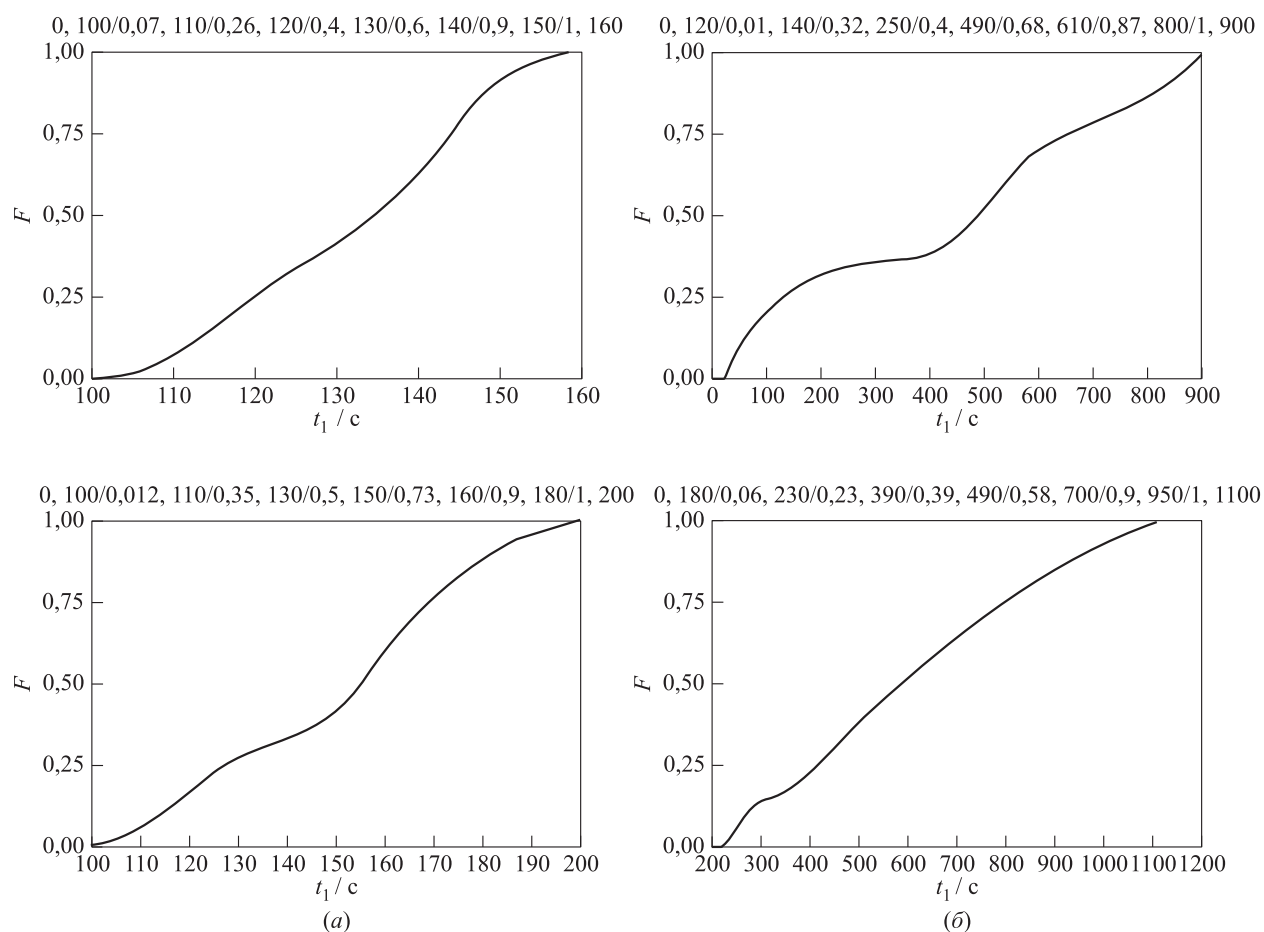


Рис. 3 Время написания первого варианта кода простой (а) и сложной (б) модели успевающим (верхний ряд) и неуспевающим учащимся (нижний ряд)

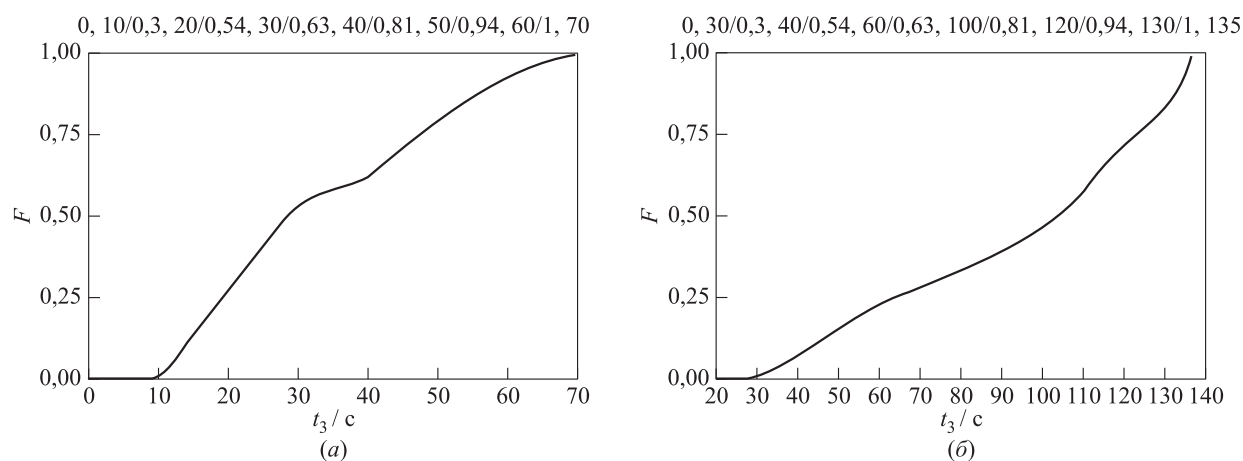


Рис. 4 Время поиска ошибок компиляции успевающим (а) и неуспевающим (б) учащимся

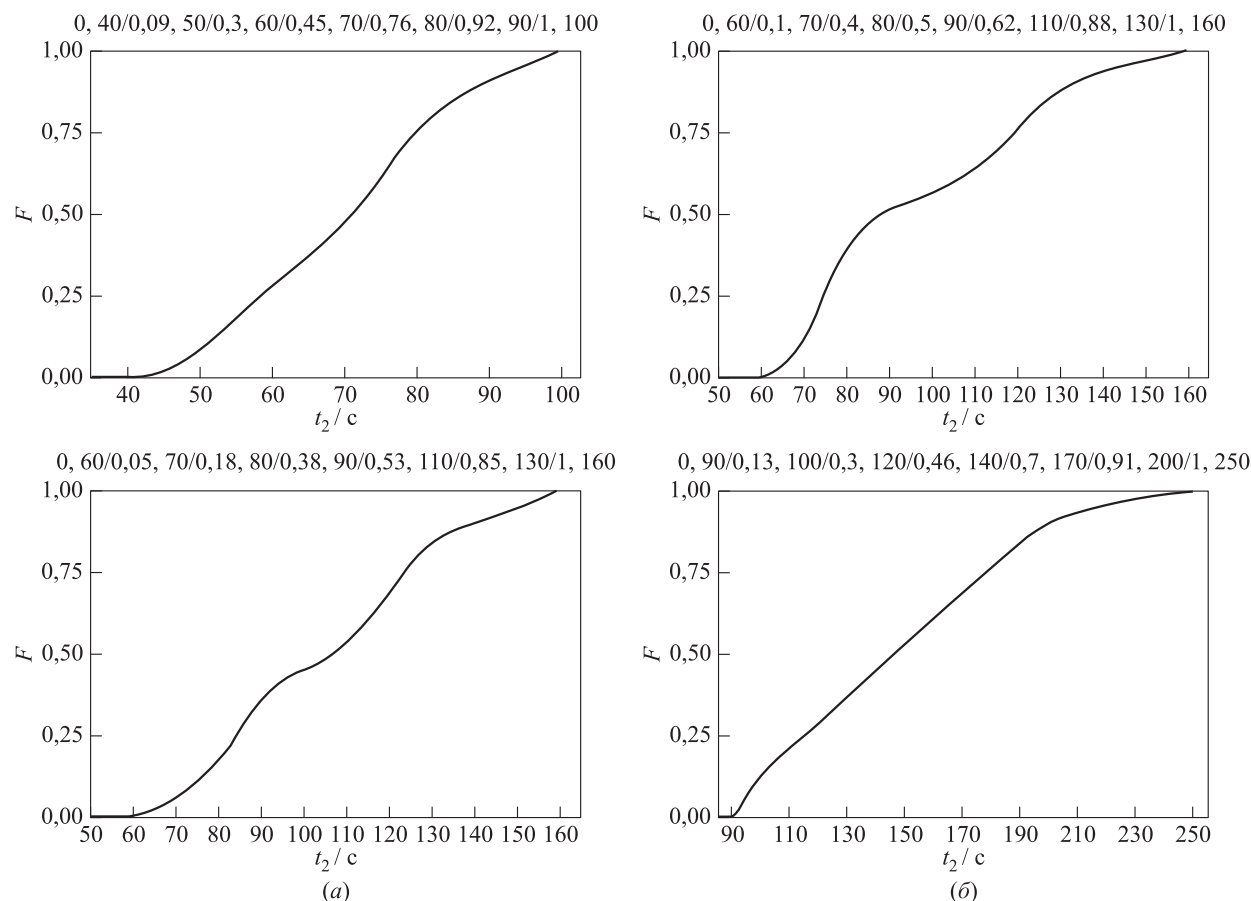


Рис. 5 Анализ отчетов простой (а) и сложной (б) модели успевающим (верхний ряд) и неуспевающим учащимся (нижний ряд)

также с периодами поиска, устранения ошибок и анализа отчета и составляет менее 0,1%.

2. Время прогона одной учебной модели варьируется в интервале [0,3–4] с в зависимости от алгоритмических свойств модели и от аппаратного ресурса, что также пренебрежимо мало в случае однопользовательского режима.

5 Сравнение отклика имитационной модели с реальной системой

Для проведения экспериментальных исследований была разработана система моделирования работы комплекса виртуальных лабораторий Open Virtual Research Space (OVRs), которая представляет собой клиент-серверное приложение для исследования процессов функционирования открытого виртуального исследовательского пространства (ОВИП) [7, 8].

Для оценки достоверности полученной модели был проведен хронометраж запросов на запуск имитационного эксперимента в среде GPSS тремя группами учащихся (табл. 2).

Таблица 2 Хронометраж запросов на запуск имитационного эксперимента

№ группы	Учащиеся		Задачи	
	Успевающие	Неуспевающие	Сложные	Простые
1	8	12	5	3
2	10	0	6	2
3	0	12	4	1

На рис. 6 приведены гистограммы пиковой загрузки реального и модельного сервера запросами на запуск имитационной модели.

Построены доверительные интервалы Велча для среднего значения числа заявок, поступающих в интервалы времени, равные 60 с: $(-0,21, 0,94)$, $(-0,83, 1,07)$; $(-1,17, 1,35)$.

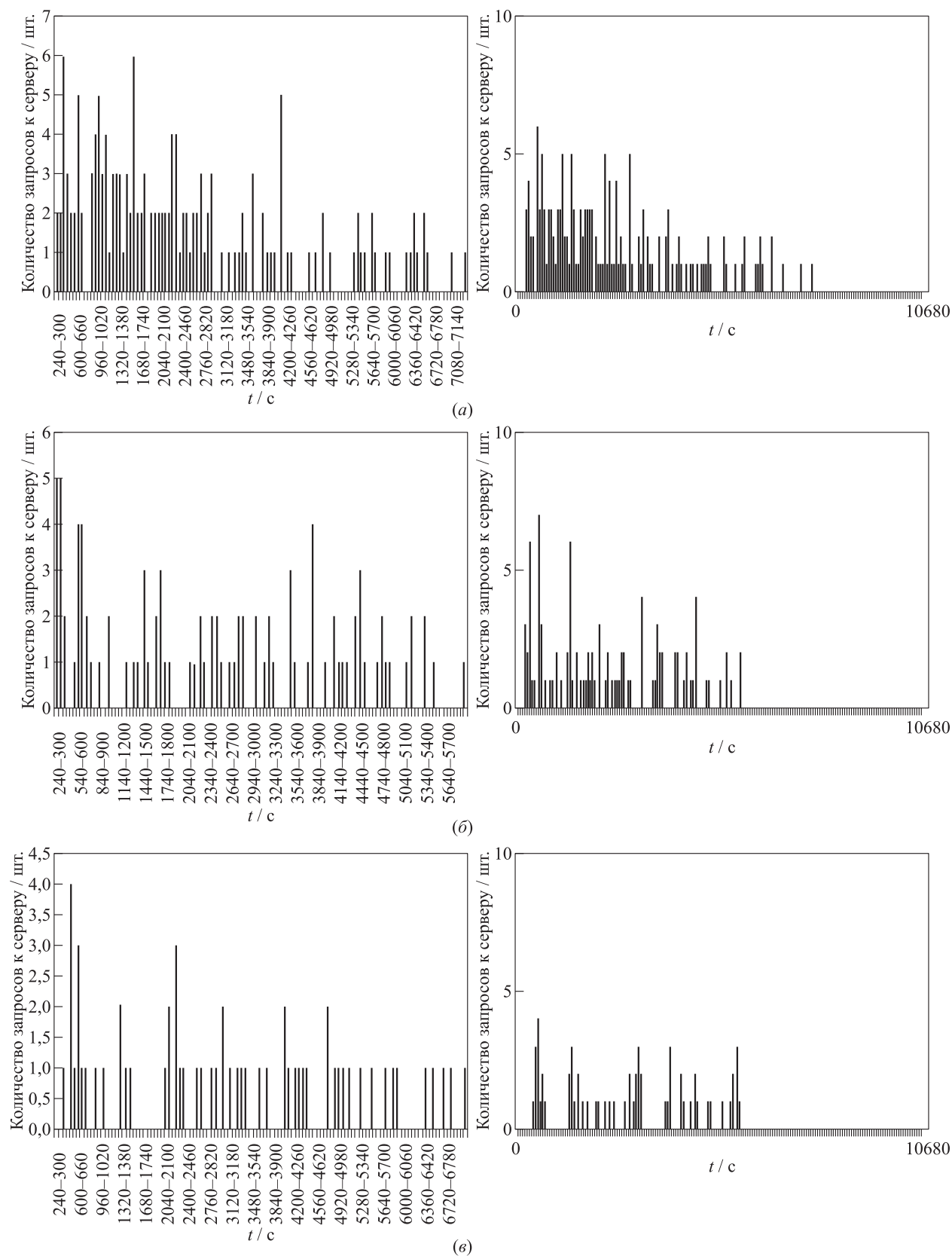


Рис. 6 Пиковая нагрузка модельного (слева) и реального (справа) сервера по табл. 2: (а) для группы 1; (б) для группы 2; (в) для группы 3

Таким образом, достоверность построенной модели подтверждается наличием нуля в каждом интервале.

6 Заключение

В ходе исследования была разработана методика оценки нагрузки на серверную часть ОСОВ, особенностью которой является моделирование потока запросов на основе системного анализа поведения групп пользователей при решении определенных задач с применением программного обеспечения.

С использованием методики были построены имитационные модели и экспериментальная среда исследования взаимодействия пользователя с системой имитационного моделирования в режиме обучения.

Построены доверительные интервалы Велча для среднего значения числа запросов, поступающих на реальный сервер и его модель в интервалы времени, равные 60 с: $(-0,21, 0,94)$, $(-0,83, 1,07)$; $(-1,17, 1,35)$, что подтверждает достоверность построенной модели.

Использование методов системного анализа для исследования поведения групп пользователей при решении определенных задач с применением средств автоматизации позволяет адаптировать предложенную методику для изучения серверной нагрузки в ОСОВ.

Методика может быть применена при построении моделей процессов взаимодействия пользователей с произвольным программным обеспечением, доступ к которому предоставлен системами облачных вычислений. При этом модели будут отражать свойства потока запросов к серверной части, приводящих к загрузке ресурсов.

Перспективным направлением развития данного исследования является теория проектирования систем *cloud computing* и, в частности, вопросы зависимости пиковых нагрузок от поведения пользователей.

Литература

1. Willinger W., Taqqu M. S., Erramilli A. A Bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks // Stochastic networks: Theory and applications. — Oxford University Press, 1996. P. 282–296.
2. Столлингс В. Современные компьютерные сети. — СПб.: Питер, 2003. 782 с.
3. Petroff V. Self-similar network traffic: From chaos and fractals to forecasting and QoS // NEW2AN. — St. Petersburg, 2004. P. 110–118.
4. Шелухин О. И., Тенякшев А. М., Осин А. В. Фрактальные процессы в телекоммуникациях. — М.: Радиотехника, 2003. 480 с.
5. Dang T. D., Sonkoly B., Molnar S. Fractal analysis and modelling of VoIP traffic // 11th Telecommunications Network Strategy and Planning Symposium (International) (NETWORKS 2004) Proceedings. — Vienna, Austria, 2004. P. 123–130.
6. Криштофович А. Ю. Применение модели трафика сети ОКС № 7 для управления потоками сигнальной нагрузки // Инфокоммуникационные технологии, 2004. Т. 2. № 2. С. 25–27.
7. Жевнерчук Д. В., Николаев А. В. Открытый инструмент проведения дистанционного имитационного эксперимента // Вестник ИжГТУ. — Ижевск: ИжГТУ, 2008. № 2. С. 103–108.
8. Ефимов И. Н., Жевнерчук Д. В., Николаев А. В. Открытые виртуальные исследовательские пространства. Аналитический обзор. — Екатеринбург: Институт экономики УрО РАН, 2008. 83 с.