

V.A. Trapeznikov Institute of Control Sciences
Academy of Sciences
RUDN University
Tomsk State University
Institute of Information and Communication
Technologies Bulgarian Academy of Sciences
Research and development company
“Information and networking technologies”

**DISTRIBUTED COMPUTER
AND COMMUNICATION NETWORKS:
CONTROL, COMPUTATION,
COMMUNICATIONS
(DCCN-2016)**

**Volume 3
Youth School-Seminar**

**Proceedings
of the Nineteenth International Scientific Conference**

Russia, Moscow, 21–25 November 2016

*Under the general editorship
of D.Sc. V. M. Vishnevskiy and D.Sc. K. E. Samouylov*

**Moscow
2016**

Федеральное государственное бюджетное учреждение науки
Институт проблем управления им. В. А. Трапезникова
РОССИЙСКОЙ АКАДЕМИИ НАУК
Федеральное государственное автономное
образовательное учреждение высшего образования
«РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ»
Национальный исследовательский
Томский государственный университет
Институт информационных и телекоммуникационных технологий
БОЛГАРСКОЙ АКАДЕМИИ НАУК
Научно-производственное объединение
«Информационные и сетевые технологии»

**РАСПРЕДЕЛЕННЫЕ КОМПЬЮТЕРНЫЕ
И ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ:
УПРАВЛЕНИЕ, ВЫЧИСЛЕНИЕ, СВЯЗЬ
(DCCN-2016)**

В трех томах

Том 3

Молодежная школа-семинар

Материалы

Девятнадцатой международной научной конференции

Россия, Москва, 21–25 ноября 2016 г.

*Под общей редакцией
д.т.н. В.М. Вишневого и д.т.н. К.Е. Самуйлова*

**Москва
2016**

УДК 004.7:004.4.001:621.391:007(063)

ББК 32.973.202:32.968

Р 24

Р 24 **Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь (DCCN-2016) = Distributed computer and communication networks: control, computation, communications (DCCN-2016)** : материалы Девятнадцатой международной научной конференции, 21–25 нояб. 2016 г. : в 3 т.; под общ. ред. В. М. Вишневого и К. Е. Самуйлова — М.: РУДН, 2016.

ISBN 978-5-209-07666-7

Т. 3: Молодежная школа-семинар = Youth School-Seminar. — 499 с. : ил. ISBN 978-5-209-07669-8 (т. 3)

Научная молодежная школа-семинар проводится при финансовой поддержке Российского фонда фундаментальных исследований, Проект № 16-37-10500.

В научном издании представлены материалы Девятнадцатой международной научной конференции «Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь» по следующим направлениям:

- Оптимизация архитектуры компьютерных и телекоммуникационных сетей;
- Управление в компьютерных и телекоммуникационных сетях;
- Оценка производительности и качества обслуживания в беспроводных сетях;
- Аналитическое и имитационное моделирование коммуникационных систем последующих поколений;
- Беспроводные сети 4G/5G и технологии сантиметрового и миллиметрового диапазона радиоволн;
- RFID-технологии и их применение в интеллектуальных транспортных сетях;
- Интернет вещей, носимые устройства, приложения распределенных информационных систем;
- Распределенные системы и системы облачного вычисления, анализ больших данных;
- Вероятностные и статистические модели в информационных системах;
- Теория очередей, теория надежности и их приложения;
- Математическое моделирование высокотехнологичных систем;
- Математическое моделирование и задачи управления.

Сборник материалов конференции предназначен для научных работников и специалистов в области теории и практики построения компьютерных и телекоммуникационных сетей.

Текст воспроизводится в том виде, в котором представлен авторами

Утверждено к печати Программным комитетом конференции

ISBN 978-5-209-07669-8 (т. 3)

ISBN 978-5-209-07666-7

©Коллектив авторов, 2016

©Российский университет дружбы народов, 2016

Contents

Abaev P. O., Beschastny V. A., Tsarev A. S., Samouylov K. E. Network Resource Management	9
Abaev P. O., Beschastny V. A., Samouylov K. E. Tractable distance distribution approximations for hardcore processes	17
Abrosimov L. I., Rudenkova M. A. Analysis of Throughput Wireless Media and Settings for Access Point Date Layer	26
Abrosimov L. I., Larin A. A. Minimization of Data Center Cost Creating Monitoring System with Determinate Data Flow	34
Adkhamova A. S., Skubachevskii A. L. About damping problem for control system with delay	41
Aliev T. I., Sosnin V. V. Characteristics of Priority Queues with High Utilization Parameter	45
Begishev V. O., Petrov V. I., Samuylov A. K., Moltchanov D. A., Gaidamaka Yu. V. Modeling NB-IoT technologies for the Internet of Things Applications.	56
Belyaeva J. O., Skubachevskii A. L. Stationary solutions of the Vlasov equations for a two-component high temperature plasma.	60
Blinov A. I., Sevastianov L. A., Vasilyev S. A. Transport systems analysis using neural networks	65
Bogatyrev V. A., Parshutina S. A. Efficiency of Redundant Multipath Transmission of Requests Through the Network to Destination Servers	73
Bolotova G. O., Vasilyev S. A., Udin D. N. Systems of Differential Equations of Infinite Order with Small Parameter and Countable Markov Chains	81
Borodina A. B., Efosinin D. V., Morozov E. V. Accelerated regenerative simulation of degradation process in the system with gradual and sudden failures	89
Dao T. N., Paramonov A. I. Analysis of communications network based unmanned aerial vehicles	92
Devyatkov V. V., Naung M. T. Model-oriented check the correctness of network interaction	101
Dinh T. D., Kirichek R. V. Development and research of methods of installation of wireless sensor nodes from UAV	114
Doligushev R. A., Kirichek R. V. An Overview of Possible Testing Types and Methods for the Internet of Things	122

Eferina E. G., Kuznetsova O. V., Korolkova A. V., Kulyabov D. S., Sevastianov L. A. Spinor representation of Maxwell's equations	129
Fomin M. B., Ivanov A. E. Recognition of anchor points on three-dimensional objects by stereo images in machine vision systems	137
Gerasimenko M. A. Neuron Networks	143
Gevorkyan M. N., Demidova A. V., Korolkova A. V., Kulyabov D. S., Sevastianov L. A. Implementation of the Wiener stochastic process in OpenModelica	150
Gevorkyan M. N., Demidova A. V., Korolkova A. V., Kulyabov D. S., Sevastianov L. A. About extensions programming for OpenModelica	158
Glukhov I. V., Orlov Y. N. Modeling of stock prices in the density distribution of a product of processes Weibull	165
Gorbunova A. V., Kradenyh A. A., Zaryadov I. S. The mathematical model of a cloud computing systeme	169
Gorshenin A. K., Korolev V. Yu. On noising of data to refine the output of moving separation of mixtures	176
Grebeshkov A. Yu., Zuev A. V., Kiporov D. S. Computer Simulation of Average Channel Access Delay in Cognitive Radio Network	184
Hussein O. A., Okuneva D. V. Analysis of D2D technologies impact on the operation of wireless networks	191
Izmaylova Y. The research of retrial queueing systems with exclusion of customers	198
Kalimulina E. Yu. Queueing System Convergence Rate	203
Kalinina K. A. Effective bandwidth estimation in highly critical systems	212
Kanzitdinov S. K., Vasilyev S. A. Neural networks with an infinite number of cells analysis	216
Kaspirovich I. E. Analisys of Numerical Solution Stability of Motion of Rolling Sphere on Rotating Plane	225
Kocheganov V. M., Zorine A. V. Low-priority queue and server's steady-state existence in a tandem under prolongable cyclic service	232
Kokshenev V., Mikheev P., Suschenko S., Tkachev R. Analysis of the effectiveness of forward error correction in selective mode of transport protocol	240
Kolbasova V., Lisovskaya E., Moiseeva S. Total capacity of customers in infinite-server queueing system with stationary renewal arrivalse	248

Kolechkin A. O., Vladyko A. G. Software for testing of controllers in software-defined networks	256
Kolomoitcev V. S., Bogatyrev V. A. The Fault-tolerant Structure of Multilevel Secure Access to the Resources of the Public Network	264
Korshok E. O., Vasilyev S. A. Solutions analysis of infinite order singular perturbed stochastics differential equation	272
Korzun D. G., Vdovenko A. S., Bogoiavlenskaia O. I. On Convergence of Active Control Strategies for Subscription Notification Delivery in Smart Spaces	281
Koucheryavy A., Makolkina M. A., Paramonov A. I. Applications of augmented reality traffic and quality requirements study and modeling	289
Kovalchukov R., Samuylov A. K., Moltchanov D. A., Andreev S., Samouylov K. E. The three-dimensional simulation framework for interference and SIR assessment	301
Kulik V. A., Vybornova A. I. Methods of complex testing the devices of the Internet of Things	305
Lisovskaya E., Moiseeva S., Pagano M. The total capacity of customers in the MMPP/GI/ ∞ queueing system	313
Makolkina M. A., Surodeeva E. V. Study of the interrelationship of subjective perception of the video quality and Hurst parameter of traffic	326
Morozov E. V., Potakhina L. V. Speed-up estimation of a system with random volume customers	334
Morozov E., Peshkova I., Rumyantsev A. On regenerative envelopes for high performance cluster simulation	337
Nazarov A. A., Fedorova E. Asymptotic analysis of retrial queue M/M/1 with impatient calls under the long patience time condition	342
Nazarov A. A., Broner V. I. Inventory management system with On/Off control and phase-type distribution of purchases quantity	349
Nazarov A. A., Pomortseva N. A. Asymptotic analysis of M/GI/1 retrial system with conflicts and afterservice	356
Okuneva D. V., Proshutinskiy K. S. Improving of the traffic balancing efficiency on the base of estimations of user attention concentration	363
Pirmagomedov R., Hudoev I., Shangina D. Simulation of Medical Sensor Nanonetwork Applications Traffic	371
Schetinin E. Yu., Merzlyakov V. Statistical extreme type dependence structures modeling in spatial domains	381

Serebryakova A. A., Kulik V. A., Pham V. D., Kirichek R. V. Effect of Traffic IoT on network equipment	388
Shilin P. A., Kirichek R. V. Research the possibility of using UAVs swarm for organization VANET infrastructure	394
Shklyayeva A. V., Kirichek R. V. An Overview of Possible Testing Types and Methods for the Flying Ubiquitous Sensor Networks	401
Sopin E. S., Daraseliya A. V., Yarkina N. V. On the virtual machines migration effectiveness in cloud systems	408
Sosnin V. V. Per-Packet Load Balancing of TCP Traffic for Goodput Aggregation of Communication Channels with Asymmetric Transmission Delay	412
Teltevskaya V. A., Makolkina M. A. Method for evaluating the quality of experience in augmented reality systems	419
Velieva T. R., Korolkova A. V., Kulyabov D. S., Sevastianov L. A. Hybrid simulation of active traffic management	427
Vishnevsky V. M., Ivanov R. E., Larionov A. A., Dudin M. S. Optimisation of data transmission scheduling in 5G mmWave backhaul network with STDMA	435
Yapo G. S., Milovanova T. A., Zaryadov I. S. Interval estimation of system performance with the optimal choice.	445
Zadiranova L., Melikov A., Moiseev A. Asymptotic Analysis of Queueing System with MMPP Arrivals and Feedback.	452
Zaryadov I. S., Matskevich I. A., Scherbanskaya A. A. The queueing system with general renovation and repeated service — time-probability characteristics.	458
Fedorov S. L. Kinetic approach in models of forecasting non-stationary time-series and functional calculations on them.	463
Gorbunova A. V., Zaryadov I. S., Matushenko S. I., Sopin E. S. The Estimation of Probability Characteristics of Cloud Computing Systems with Splitting of Requests.	467
Houankpo H. G. K., Kozyrev D. V. Sensitivity analysis of steady state reliability characteristics of a cold redundant data transmission system to the shapes of lifetime and repair time distributions of its elements	473
Ustyuzhanin N., Gilmutdinov M. Analysis of background modeling methods performance in lossy video compression systems	481
Zaryadov I. S., Razumchik R. V. Stationary waiting time in $G/M/n/r$ FCFS queue with random renovation	489

Author index

498

УДК 621.39

Об оценке времени пребывания в очереди в программно-конфигурируемых сетях

П. О. Абаев, В. А. Бесчастный,
А. С. Царёв, К. Е. Самуйлов

*Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклуто-Маклая, д.6, Москва, Россия, 117198*

Аннотация. Задержка в современных сетях является одной из ключевых характеристик качества обслуживания и имеет особое значение для таких приложений, для которых не допускаются паузы при передачи данных. К таким приложениям можно отнести интерактивное видео и сетевые игры. В этой работе для решения задачи обеспечения качества обслуживания (QoS, Quality of Service) по задержкам используются преимущества архитектуры SDN (Software Defined Networking). В работе производится оценка задержек в модели SDN-сети в условиях поступающего трафика, приближенного по характеристикам к реальному, а также сквозного управления качеством обслуживания. На основе построенной модели предложен метод управления общей задержкой в сети, который позволяет обеспечивать QoS-требования с помощью динамического переключения потоков данных между узлами, обладающих различными характеристиками.

1. Введение

Обеспечение качества обслуживания имеет особо важное значение для таких приложений реального времени, как видеоконференции и сетевые игры. Архитектура SDN доказала свою эффективность, позволив реализовать целый ряд приложений, имеющих активное применение в современных сетях [1, 2]. Принцип разделения сетевых функций на уровень управления и уровень данных позволяет добиться более гибкого и адаптивного управления. В частности, центральный контроллер может отслеживать состояние всей сети, что позволяет применять довольно простые методы управления. Это свойство имеет особое значение для большинства приложений, которым необходимо сохранять баланс между производительностью и алгоритмической сложностью. Интерактивные сетевые приложения реального времени требуют выполнения требований к задержкам на протяжении всего периода своего функционирования. Однако задержки в сети подвержены большим колебаниям в условиях произвольного трафика и перегрузок в сети. Цель работы состоит в оценке времени ожидания начала обслуживания и управлении общей задержкой в SDN-сети. В работе решаются следующие задачи:

- Моделирование задержки в узле сети с учетом его характеристик, таких как размер буфера, пропускная способность, номер обслуживаемого потока данных и задержка распространения сигнала, вычисляемая с помощью SDN-контроллеров;
- Расширение модели для оценки общей задержки на протяжении всего маршрута прохождения потока;
- Разработка метода управления задержками для потоков в среде SDN, состоящего из трех частей: подбор параметров сети, оценка задержки и динамическое переключение потоков между узлами сети.

2. Описание модели

В работе предполагается, что задержка передачи данных постоянна, величиной задержки обработки можно пренебречь, а время нахождения в очереди является случайной величиной и составляет основную часть общей задержки. В качестве основного показателя для анализа системы используется средняя задержка, так как с ее помощью можно получить оценки задержки, в то время как мгновенная задержка (джиттер) может быть нивелирована за счёт очередей.

В качестве модели трафика выберем ТСП-трафик, так как ТСП является наиболее популярным протоколом транспортного уровня в современных сетях. Так как в ТСП применяется управление перегрузками, то интенсивность поступления заявок отличается от модели UDP-трафика. Сначала будет рассмотрена модель с одним потоком ТСП-трафика, затем будет рассмотрен случай с несколькими потоками, значительно лучше отображающий реальные данные.

3. Преимущества модели на основе SDN

С точки зрения выбора параметров, SDN может позволить построить более простую и практичную модель. Благодаря централизованной архитектуре уровня управления сети, можно получить такие параметры, как количество потоков, топологию сети, состояния очередей и задержку передачи данных. Учитывая нечастые изменения параметров, за исключением количества потоков, будем измерять эти параметры сразу после инициализации сети, а также после реконфигурации и событий, влияющих на состояние сети. Таким образом, задача измерения параметров становится довольно простой. В качестве SDN контроллера был выбран Floodlight, написанный на языке программирования Java, который является свободно распространяемым программным обеспечением и поддерживает протокол OpenFlow.

С точки зрения выбора параметров, SDN может позволить построить более простую и практичную модель. Благодаря централизованной архитектуре уровня управления сети, можно получить такие параметры, как количество потоков, топологию сети, состояния очередей и задержку передачи данных. Учитывая нечастые изменения параметров, за исключением количества потоков, будем измерять эти параметры сразу после инициализации сети, а также после реконфигурации и событий, влияющих на состояние сети. Таким образом, задача измерения параметров становится довольно простой. В качестве SDN контроллера был выбран Floodlight, написанный на языке программирования Java, который является свободно распространяемым программным обеспечением и поддерживает протокол OpenFlow.

Задержка передачи данных - это один из параметров, используемых для оценки общей задержки, представляющий время прохождения сигнала в сети без учета задержки в очереди. Задержку передачи достаточно измерить только один раз, когда нагрузка в сети низкая. Поскольку OpenFlow имеет возможности для отправки пакетов в сеть, мы можем получить канальную задержку для OpenFlow, как показано в [5,6]. Если нагрузка на сеть низкая, то канальная задержка равняется задержке передачи. Для того, чтобы из канальной задержки получить задержку передачи на протяжении всего пути, необходимо отправить пакет с первого коммутатора S_1 на последний коммутатор S_m . Перед измерением данных отправленного пакета, необходимо настроить элементы потока согласно правилам перенаправления. На первом шаге контроллер, инкапсулировав все необходимые поля пакета, отправит его на коммутатор S_1 . Затем S_1 направит его своему коммутатору-соседу S_2 , далее это шаг будет повторяться, пока пакет не попадет на последний коммутатор в пути S_m . На последнем шаге пакет будет возвращен обратно на контроллер с коммутатора S_m . Контроллер записывает 2 отметки времени для этого пакета как время отправки $T_{\text{отправ}}$ и время возвращения $T_{\text{возвр}}$. С учетом дополнительной задержки между контроллером и коммутатором, контроллер также должен рассчитать время прохождения сигнала в обоих направлениях RTT_{C-S_1} и RTT_{C-S_m} между контроллером и двумя конечными коммутаторами используя тот же метод. Общая задержка распространения T_q может быть выражена как (1).

$$T_q = T_{\text{возвр}} - T_{\text{отправ}} - 0,5 \cdot (RTT_{C-S_1} + RTT_{C-S_m}) \quad (1)$$

Последняя версия OpenFlow не обеспечивает возможностей для управления очередью. Контроллер может запрашивать только некоторую ограниченную и статическую конфигурацию параметров посредством OpenFlow, например, максимальную или минимальную интенсивность, но создавать или изменять очередь протокол не может. Однако, в некоторых работах были рассмотрены альтернативные методы управления очередями в системах, базирующихся на OpenFlow. В данной модели применимы два метода, описанные ниже.

В [4] описан метод управления правилами планировщика очереди посредством отправки и запроса информации о пространстве ядра контроллера. Данный метод основан на Netlink, который используется в качестве IPC (межпроцессного взаимодействия) между пользовательском пространством Linux и пространством ядра. Сообщения обеспечения QoS от контроллера отображаются в виде Netlink-сообщения, которое затем посылается подсистеме Linux по управлению трафиком ядра.

Кроме того, существует еще один способ расширить протокол конфигурации OpenFlow (OF-CONFIG) для обеспечения возможности управления очередью. Задача OF-CONFIG заключается в удаленной настройке путей потока данных в OpenFlow. Рассматриваемый метод основывается на модели данных Netconf и Yang [3]. Точка конфигурации OpenFlow может быть сервисом, предоставляемым программой, такой как OpenFlow контроллер, либо она может быть сервисом, предоставляемым традиционными средами управления сетью.

4. Основные обозначения и параметры модели

Протокол OpenFlow позволяет отслеживать трафик (т.е. тип протокола и номер потока) в каждом соединении. Контроллер по запросу может собирать статистику каждого потока. Количество потоков может быть рассчитано в момент поступления нового потока на контроллер.

В среде SDN, конечные узлы подключены через SDN OpenFlow коммутаторы, контролируемые Floodlight контроллером. Пусть N TCP потоков отправляются с узла H_1 на узел H_2 . W - размер окна потока. W_{min} , W_{max} и W_{cp} представляют минимальный, максимальный и средний размер окна соответственно. На каждом коммутаторе создается несколько очередей с накопителем размера B и обслуживающей способностью C . Потоки будут обслуживаться в разных очередях в зависимости от требований к качеству обслуживания. Если интенсивность поступления заявок выше пропускной способности, размер очереди L_q увеличивается. Контроллер имеет доступ к информации о размере накопителя и обслуживающей способности посредством запросов к коммутаторам. Кроме информации об очереди, также необходима информация о задержке передачи на пути T_p и числе потоков в очереди, чтобы оценить задержку нахождения в очереди T_q . Таким образом мы получим конечную задержку на всем пути $T = T_p + T_q$. Необходимо оценить среднюю задержку пути (2) на основе параметров сети.

$$T_{cp} = T_p + T_{q_cp} \quad (2)$$

5. Оценка средней задержки в случае одного ТСП-потока

Как для случая одиночного ТСП потока, так и для нескольких потоков, задержка нахождения в очереди напрямую зависит от интенсивности отправки сообщений (размера окна). Нагрузка на ТСП сеть из-за размера окна увеличивается линейно и уменьшается в половину, когда происходит потеря пакета, и, в свою очередь, влияет на длину очереди и задержки нахождения в очереди. После достаточно большого количества наблюдений, можно утверждать, что размер окна одиночного потока ТСП имеет равномерное распределение.

Учитывая, что T_{q_cp} в (2) можно получить из (3), где L_{q_cp} - это средняя длина очереди, $L_{пкт}$ - размер пакета (в байтах). Добавим множитель 8 для преобразования $L_{пкт}$ из байт в бит, а C - пропускная способность, получим:

$$T_{q_cp} = \frac{L_{q_cp} \cdot L_{пкт} \cdot 8}{C} \quad (3)$$

Неподтвержденные пакеты будут распределены в соединении очереди, а связь между L_q и W можно выразить как (4). L_{q_cp} можно рассчитать как W_{cp} со смещением:

$$L_q = W - \left(\frac{2 \cdot T_p \cdot C}{8 \cdot L_{пкт}} \right) \quad (4)$$

Из-за линейного роста размера окна, $W_{cp} = (W_{\min} + W_{\max})/2$, окно ТСП уменьшается наполовину, достигая значения $W_{\min} = W_{\max}/2$, тогда $W_{cp} = 3W_{\max}/4$. В случае максимального использования канала, когда очередь будет полностью занята, получим (5):

$$W_{max} = B_0 + \left(\frac{2 \cdot T_p \cdot C}{8 \cdot L_{пкт}} \right) \quad (5)$$

Средняя задержка T_{cp} одиночного ТСП потока под B_0 выражена в (7):

$$T_{cp} = \frac{1}{2}T_p + \frac{3}{4} \cdot \frac{B \cdot L_{пкт} \cdot 8}{C} \quad (6)$$

$$T_{cp} = T_{cp} + 0.002 \cdot \frac{B_1}{B_0} \quad (7)$$

6. Оценка средней задержки в случае нескольких ТСП-потоков

В случае нескольких несинхронизированных потоков, время нахождения в очереди отображает совершенно другую характеристику, в отличие от случая с одним потоком. В данном случае, так как

несколько потоков занимают одну очередь, размер очереди связан с совокупным размером окна. При увеличении числа потоков, размер окна различных потоков оказывается десинхронизирован. Сначала мы рассмотрим распределение агрегированного размера окна, а затем время нахождения в очереди.

При увеличении числа потоков, распределение агрегированных размеров окна будет изменяться в сторону большего среднего значения и меньшего отклонения. Это происходит потому, что изменение агрегированного размера ТСР окна уменьшается с увеличением количества помех между различными потоками в одной общей очереди. В [7] установлено, что в случае нескольких ТСР потоков каждый из потоков соответствует гамма-распределению, а не равномерному, упомянутому выше. Кроме того, совокупный размер окна из нескольких ТСР потоков принимает вид нормального распределения [8] при одновременном числе соединений более 1000. Однако на данный момент установлено, что агрегированный размер окна показывает хорошие показатели десинхронизации при количестве подключений больше 10. В данной работе предполагается, что нормальное распределение соответствует как малому, так и большому количеству потоков.

В случае нескольких потоков, размер очереди имеет то же распределение, что и агрегированной размер окна.

Средняя задержка по маршруту для N потоков состоит из задержки передачи и среднего времени нахождения в очереди, и может быть выражена как (8):

$$T_{cp}(N) = T_p + \frac{\mu \cdot L_{пкт} \cdot 8}{C} \quad (8)$$

Согласно результатам из [8], теоретический размер очереди для нескольких потоков выражен в (9):

$$B_0 = \frac{2 \cdot T_p \cdot C}{8 \cdot L_{cp} \cdot \sqrt{N}} \quad (9)$$

Приняв B_1 за установленный размер очереди, это значение может не удовлетворять B_0 для нескольких потоков. Если мы предположим, что α - коэффициент соответствия размера очереди, основанный на функции ошибки нормального распределения размера очереди, то в (10) можно получить ограничения вероятности использования при $\alpha < 1$:

$$\min \{(PX \leq B_i)\} = erf \left(\frac{3\sqrt{3}}{2\sqrt{2}} \cdot \frac{\alpha}{1 + \frac{\alpha}{\sqrt{N}}} \right) \quad (10)$$

Из теории о нормальном распределении имеем $\varphi(\alpha) = P\{X \leq (\mu + \alpha \cdot \sigma)\}$, где μ - среднее значение, а σ - отклонение нормального распределения (11):

$$\sigma = \frac{2 \cdot T_p \cdot C}{8 \cdot L_{\text{пкт}}} + B_1 \quad (11)$$

$$3 \cdot \sqrt{3} \cdot \sqrt{N}$$

Пусть дано $\min \{P(X \leq B_i)\} = \varphi(\alpha)$, тогда коэффициент α в (10) может быть отыскан. Предположим, что B_1 эквивалентно $(\mu + \alpha \cdot \sigma)$, тогда $\mu = B_1 - \alpha \cdot \sigma$, и соответственно средняя конечная задержка $T_{\text{ср}}(N)$ для N потоков может быть получена как (12):

$$T_{\text{ср}}(N) = T_p + \frac{8 \cdot B_1 \cdot L_{\text{пкт}}}{C} - \alpha \cdot \frac{2 \cdot T_p \cdot C + B_1 \cdot L_{\text{пкт}} \cdot 8}{3\sqrt{3} \cdot \sqrt{N}} \quad (12)$$

Благодарности

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

Литература

1. *Agarwal S., Kodialam M.* Traffic engineering in software defined networks // Proceedings IEEE INFO.COM, vol 12, no11, — P. 2211-2219, 2013
2. *Jain S.* Experience with a Globally-Deployed Software Defined WAN // Acm Sigcomm Computer Communication Review, vol 43, no4, — P. 3-14, 2013
3. *Fgee B., Kenney J.D., Phillips W.J., et al* Comparison of QOS Performance between IPV6 QOS Management Model and Int Sery and DifServ Qos Models // Communication Networks and Services Research, Annual Conference on IEEE Computer Society, — P. 287-292, 2005
4. *Ishimori A., Farias F.* Control of multiple packet schedulers for improving QoS on OpenFlow/SDN networking // Software Defined Networks, EWSDN 2013, — P. 81-86, 2013
5. *Phemius K., Bouet M.* Monitoring latency with Open Flow // 9th International Conference on Network and Service Management (CNSM) IEEE Computer Society, — P. 122-125, 2013
6. *Van Adrichem* Open Net Mon Network monitoring in Open Flow Software-Defined Networks // Network Operations and Management Symposium (NOMS), — P. 1-8, 2014
7. *Yan J., Plattner B.* A Simple Solution to Find the Distribution of TCP Window Sizes // IEEE Communications Letters, vol. 17, — P. 417-419, 2013
8. *Appenzeller G., Keslassy I., Mc Keown N.* Sizing router buffers // Acm Sigcomm Computer Communication Review, vol. 34, no4, — P. 281-292, 2004

UDC 621.39

Network Resource Management

**P. O. Abaev, V. A. Beschastny, A. S. Tsarev,
K. Ye. Samouylov**

*Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

УДК 621.39

Аппроксимация распределений расстояний в процессах твёрдого ядра

П. О. Абаев, В. А. Бесчастный, К. Е. Самуйлов

*Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Макляя, д.6, Москва, Россия, 117198*

Аннотация. Пуассоновский точечный процесс (ПТП), основным преимуществом которого является относительная легкость аналитических методов его анализа, являющаяся следствием существования распределений расстояний до n -соседа в явном виде, широко применяется для анализа производительности беспроводных сетей.

В то же время в ПТП допускается сколь угодно малое расстояние между взаимодействующими устройствами, что выражается не только в расхождении с реальными данными, но и в сложности анализа с использованием моделей распространения сигнала, включающих степенные зависимости характеристик. В качестве альтернативного подхода часто применяют процессы твёрдого ядра, в которых всегда соблюдается определённая минимальная дистанция между точками процесса, благодаря чему такие процессы лишены недостатков, присущих ПТП. Однако для процессов твёрдого ядра не существует аналитических выражений расчёта распределений расстояний. В работе исследованы распределения расстояний в процессе твёрдого ядра Матерна, а также предложен метод аппроксимации на основе алгоритма EM-типа. Представлены результаты расчета параметров аппроксимирующих гамма-распределений для различных интенсивностей исследуемого процесса.

Ключевые слова: процесс твёрдого ядра, аппроксимация, пуассоновский процесс, распределение расстояний, SINR.

1. Введение

Стохастическая геометрия не так давно стала использоваться специалистами по беспроводным технологиям в качестве инструмента изучения характеристик производительности современных и перспективных беспроводных технологий [2, 4, 5]. Первым этапом при моделировании большинства беспроводных систем является выбор пространственного процесса, определяющего местоположение узлов сети. Во многом благодаря тому, что предсказать размещение пользователей на плоскости заранее не представляется возможным, для моделирования их расположения, как правило, используются случайные пространственные процессы. В зависимости от конкретных целей выбор может пасть на один из процессов, таких как: пуассоновский пространственный процесс, кластерный пуассоновский процесс, биномиальный пространственный процесс, процесс твердого ядра и др [3].

Соотношение сигнал-интерференция плюс шум является фундаментальным показателем, характеризующим производительность мобильных беспроводных систем [6]. Рассчитав SINR можно перейти к расчету теоретически максимально возможной скорости передачи данных, которая может быть выражена как ёмкость канала по Шеннону $C = B \log_2(1 + S)$, где S представляет SINR. Далее, задав тип модуляции и схемы кодирования сигнала, с помощью SINR можно рассчитать спектральную эффективность и ёмкость сегмента сети на базе произвольной беспроводной технологии. Формально SINR задается формулой $S(\vec{d}, \vec{P}_T, f) = \frac{P_{R_0}(d_0, P_T, f)}{N_0 + I(\vec{P}_T, \vec{d}, f)}$, где $P_{R_0}(d_0, P_T, f)$ - спектральная плотность принимаемого сигнала на расстоянии d_0 , $I(\vec{P}_T, \vec{d}, f)$ - суммарная спектральная плотность сигнала (СПС) от интерферирующих источников на приемнике, N_0 - СПС шума на приемнике, \vec{d} - вектор расстояний от интерферирующих источников до приемника, f - рабочая частота.

Для моделирования интерференции и SINR, как правило, используются пуассоновские пространственные процессы. Причина в том, что распределения расстояний могут быть выражены аналитически в виде обобщенных гамма-распределений [1]. Однако при их использовании возникает следующая сложность: они позволяют расстояниям между точками быть бесконечно малыми. Это свойство не позволяет им соответствовать реальности, поскольку между пользователями сети всегда присутствует некоторая дистанция. К тому же, оно приводит к тому, что среднее значение интерференции не всегда существует: интеграл $E[I] = \int_0^\infty f(x)1/x^{-\gamma}dx$, где $f(x)$ – расстояние до ближайшего соседа, γ – коэффициент затухания сигнала в среде, расходится, что явно отличается от действительности.

Более правдоподобную оценку можно получить с помощью пуассоновского процесс твёрдого ядра, также называемого процессом твёрдого ядра Матерна (англ. Matern Hardcore Process, МНС), который может быть получен с помощью специальной процедуры «просеивания» («прореживания») исходного пуассоновского процесса [8]. Отметим, что для МНС-процессов не существует выражений, позволяющих вычислять распределения расстояний до ближайших соседей, что делает невозможным применение аналитических методов исследования беспроводных сетей. В работе исследуются распределения расстояний в МНС-процессе, а также показано, что аппроксимация на основе прореянного ПТП приводит к значительной потере точности расчетов как в отношении распределений расстояний, так и в отношении интерференции, особенно для нескольких ближайших соседей, которые вносят наибольший вклад в значение интерференции. Далее предложена аппроксимация этих распределений расстояний с использованием алгоритма ЕМ-типа. Представлены параметры аппроксимирующих гамма-распределений расстояний до нескольких ближайших соседей для различных интенсивностей МНС-процесса.

2. Пуассоновские точечные процессы и процессы твёрдого ядра

Пуассоновские точечные процессы (англ. Poisson Point Process, PPP) используется для моделирования или абстрактного представления сети, состоящей из, возможно, бесконечного числа узлов, произвольно и независимо распределенных на конечной или бесконечной зоне покрытия. В процессе твердого ядра Матерна вводится условие минимальной дистанции r между любыми двумя точками процесса. Он может быть получен с помощью зависимого просеивания PPP: в порождающем PPP каждой точке присваивается случайная метка m_i равномерно распределенная на $[0, 1]$, затем удаляются все точки, находящиеся на расстоянии менее r от точки с меньшим значением метки. Таким образом, остаются только точки, имеющие наименьшую метку в своем r -окружении. В результате в построенном процессе твёрдого ядра не найдется ни одной пары точек с расстоянием менее r между ними.

Пуассоновским точечным процессом называется точечный процесс $\Pi = \{x_i; i = 1, 2, 3, \dots\} \subset \mathbb{R}^d$ тогда и только тогда, когда число точек в произвольном компактном подмножестве $B \subset \mathbb{R}^d$ является случайной величиной с распределением Пуассона, а количество точек в подмножестве не зависит от количества в другом произвольном непересекающемся подмножестве. Другими словами: Π является PPP тогда и только тогда, когда для любого $k = 1, 2, 3, \dots$ взаимно непересекающиеся $A_i \subset \mathbb{R}^d$, где $i = 1, \dots, k$, $(\Pi(A_1), \dots, \Pi(A_k))$ является вектором независимых случайных велчин, имеющих пуассоновские распределения с параметрами $\Lambda(A_1), \dots, \Lambda(A_k)$ соответственно.

Процесс твердого ядра - это точечный процесс, в котором любые две точки расположены друг от друга на расстоянии не менее заданного параметра твердого ядра r . Точечный процесс $\Pi = \{x_i; i = 1, 2, 3, \dots\} \subset \mathbb{R}^d$ является процессом твёрдого ядра тогда и только тогда, когда в $\Pi = \{x_i; i = 1, 2, 3, \dots\} \subset \mathbb{R}^d$ $\|x_i - x_j\| \geq r$ $\forall x_i, x_j \in \Pi, j \neq i$, где $r \geq 0$ предопределённый параметр твёрдого ядра.

В [1] представлены выражения для расчёта распределений расстояний от произвольной точки плоскости до n -ближайшего соседа в виде гамма-распределений (1):

$$f(r, n) dr = \frac{2(\pi\lambda)^n}{(n-1)!} r^{2n-1} e^{-\lambda\pi r^2}, r > 0, n = 1, 2, \dots \quad (1)$$

Процесс твёрдого ядра может быть получен из порождающего PPP с помощью процедуры просеивания. В результате просеивания получается новый процесс меньшей либо равной интенсивности. Такой процесс называется процессом твёрдого ядра Матерна. Поскольку просеивание является зависимым (точки удаляются в зависимости от расстояния между ними), МНС не обладает свойством независимости, что приводит к невозможности вывода аналитических выражений.

Сначала на шагах 1-4 производится построение порождающего PPP с интенсивностью $\Lambda = \lambda \cdot S$, где S – площадь рассматриваемой плоскости (рис. 1.а). Также на шаге 3 каждой точке присваивается случайная метка m_i . Затем рассчитываются расстояния между всеми парами точек. Если расстояние меньше предустановленного параметра r , то точка с меньшим значением метки остается, а другая удаляется. После того как все пары точек обработаны, в результате получается МНС (рис. 1.б).

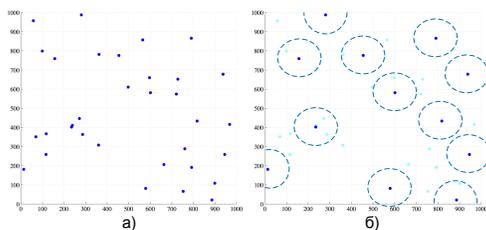


Рис. 1. Порождающий PPP (а) и полученный МНС-процесс (б).

Несмотря на то, что проблема бесконечно малых расстояний между точками в PPP известна уже на протяжении многих лет, пуассоновские распределения использовались в ряде фундаментальных исследований производительности мобильных беспроводных сетей. В некоторых из них предложены непрямые методы, позволяющие от неё отойти: например, в [7, 8] авторы предполагают, что вокруг приемника всегда найдется достаточно малая область радиуса r , вероятность попадания точки PPP в которую будет ничтожно малой так, что ей можно пренебречь. Однако они также отмечают, что выбор r является случайным, что может повлиять на итоговое значение SINR в системе. Для анализа расстояний между узлами сети PPP с просеянной интенсивностью дает довольно неточную аппроксимацию [8].

3. Анализ распределений расстояний

В этом разделе проводится сравнительный анализ распределений расстояний в PPP и МНС. Согласно [9], устройства, расположенные

на расстоянии более чем $2r$ от приемника, оказывают одинаковую интерференцию в обоих типах процессов. По этой причине в работе рассматриваются расстояния от целевого передатчика до ближайших интерферирующих. Кроме того, без учета мощностей передачи сигнала интерферирующих устройств чем ближе расположено устройство от приемника, тем больше оно оказывает влияние на передачу сигнала. На рис. 2 представлены распределения расстояний до трех ближайших соседей в PPP и МНС одинаковой интенсивности.

Принципиальное отличие между PPP и МНС заключается в том, что в PPP расстояние между целевым и интерферирующими источниками может бесконечно малым в независимости от параметров сети, что приводит практически к обнулению значения SINR и не соответствует действительности. Более того, в PPP число точек ничем не ограничено и может быть сколь угодно большим. В процессах твёрдого ядра, напротив, всегда существует максимальное число точек, которые могут быть распределены по плоскости конечной площади. Стоит отметить, что точки в процессе твердого ядра по сравнению с PPP располагаются плотнее к окружности радиуса r с центром, соответствующим положению целевого приемника (рис. 2).

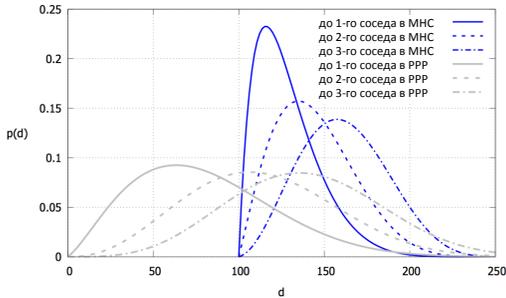


Рис. 2. Распределения расстояний до трёх ближайших соседей в PPP и МНС.

Очевидно, что погрешность измерений с использованием PPP в большинстве случаев окажется недопустимой, особенно при расчёте SINR. Поскольку для МНС не существует аналитических выражений, необходима приемлемая аппроксимация распределений расстояний до ближайших соседей. Так как точки за радиусом $2r$ в обоих типах процессов привносят одинаковый вклад в значение интерференции, далее аппроксимация будет производиться только для точек в радиусе $2r$ от

целевого приемника. Следовательно, необходимо определить количество ближайших соседей N , расстояния до которых будут аппроксимироваться. Из соображений здравого смысла, число N не может превышать девяти, поскольку невозможно вписать более девяти окружностей радиуса r в окружность радиуса $(2r + r)$, где $2r$ соответствует радиусу аппроксимируемой области, а r – радиусу окружности, центр которой может попасть на её границу. Таким образом, девяти расстояний всегда достаточно, чтобы получить точную аппроксимацию МНС. Для сокращения вычислительной сложности предлагается использовать (2), где φ – так называемый «уровень значимости».

$$N = \inf \mathbb{N} \left\{ \forall \eta, \eta \in \mathbb{N} \Rightarrow \int_{2r}^{\infty} f(\rho, \eta) d\rho < \varphi \right\}, \quad (2)$$

где $f(\rho, \eta) = \frac{2(\pi\lambda_{\text{МНС}})^\eta}{(\eta-1)!} \rho^{2\eta-1} e^{-\lambda_{\text{МНС}}\pi\rho^2}$.

Чем меньше значение φ , тем больше точность и вычислительная сложность расчетов. В этой работе рассматривается зона обслуживания в виде плоскости квадратной формы со стороной 10^3 метров. Решение неравенства (2), где $\varphi = 0.05$, можно представить в следующем виде: $N = 6$ при $\Lambda = 10^{-4} : 5.5 \times 10^{-4}$, и $N = 7$ при $\Lambda = 5.5 \times 10^{-4} : 10^{-3}$. Для меньших интенсивностей аппроксимация не имеет смысла, так как МНС, состоящий всего из нескольких точек достаточно хорошо аппроксимируется PPP той же интенсивности.

4. Метод аппроксимации и численный анализ

В этом разделе предложен метод аппроксимации распределений расстояний в МНС-процессе. Приемлемая аппроксимация должна удовлетворять двум требованиям: (i) модель должна позволять аппроксимировать эмпирические плотности с любой заданной точностью, и (ii) аппроксимация должна позволять проводить аналитический анализ беспроводных коммуникационных систем. В работе предложен метод аппроксимации, который позволяет построить процесс, подобный процессу твёрдого ядра Матерна. Метод включает аппроксимацию N -ближайших соседей и использует алгоритм EM-типа для расчета параметров гамма-распределений. Более конкретно для заданной интенсивности аппроксимируемого МНС-процесса вычисляется интенсивность порождающего PPP-процесса по формуле $\lambda_{\text{PPP}} = \frac{\ln(1 - \lambda_{\text{МНС}}\pi r^2)}{\pi r^2}$.

Затем рассчитываются эмпирические распределения расстояний до N -ближайших соседей, из которых вычитается параметр r . Далее с помощью EM-алгоритма вычисляются параметры аппроксимирующих гамма-распределений. В результате получаются N пар параметров (k, θ) .

Результирующий аппроксимирующий точечный процесс может быть построен следующим образом: генерируется число $N_{AP} \sim Pois(\Lambda)$ точек на плоскости, из которых N точек распределены по найденным гамма-распределениям с добавлением сдвига r . Остальные $(N_{AP} - N)$ точек моделируются с помощью выражения (1). Точность аппроксимации на примере SINR проиллюстрирована на рис. 3.

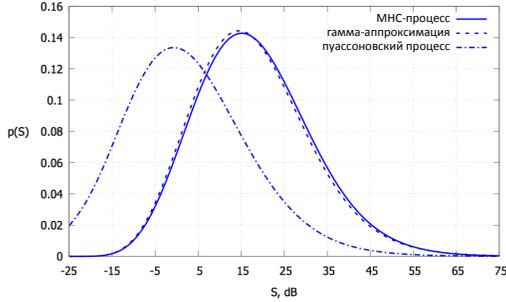


Рис. 3. Распределение SINR и аппроксимирующих распределений.

5. Заключение

В работе проведен анализ распределений расстояний в МНС-процессе а также показано, что для любой интенсивности МНС-процесса всегда существует некоторый номер N , начиная с которого распределения расстояний до ближайших соседей в просеянном PPP и МНС идентичны с точки зрения расчета SINR. Для распределений расстояний до N -первых соседей предложен метод аппроксимации на основе гамма-распределений. Важным отличительным свойством предложенного метода аппроксимации является возможность применения аналитических методов анализа беспроводных технологий, что позволит решать задачи нелинейной оптимизации, которые возникают при планировании сети, например, задачи по повышению энергоэффективности системы.

Благодарности

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

Литература

1. *Moltchanov D.* Distance distributions in random networks //Elsevier Ad Hoc Networks — Vol. 10, no. 6. — P. 1146–1166, 2012.
2. *ElSawy H., Hossain E. and Haenggi M.* Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey //IEEE Comm. Sur. Tut. — Vol. 15, no. 3. — P. 996–1019, 2013.
3. *Cox D. and Isham V.* Point processes. — Chapman and Hall/CRC Monographs on Statistics and Applied Probability, 1980.
4. *Andrews J., Ganti K., Haenggi M., Jindal N. and Weber S.* A primer on spatial modeling and analysis in wireless networks //IEEE Comm. Sur. Tut. — Vol. 48, no. 11. — P. 156–163, 2010.
5. *Haenggi M, Andrews J., Bacelli F., Dousse O. and Franceschetti M.* Stochastic geometry and random graphs for the analysis and design of wireless networks //IEEE JSAC. — Vol. 27, no. 11. — P. 1029–1046, 2009.
6. *Rappaport T.* Wireless Communications: Principles and Practice. — Prentice Hall, 2002.
7. *Petrov V., Moltchanov D. and Koucheryavy Y.* Interference and SINR in Dense Terahertz Networks. In: Proc. VTC-Fall11, — P. 1437–1442, 2015.
8. *Bacelli F. and Blaszczyszyn B.* Stochastic Geometry and Wireless Networks, Part I: Theory. — Now Publishers, 2009.
9. *Busson A. and Chelius G.* Point Process for Interference Modeling in CSMA/CA Ad-Hoc Networks //Research Report RR-6624, Inria. — P. 1–11, 2008.

UDC 621.39

Tractable distance distribution approximations for hardcore processes

P. O. Abaev, V. A. Beschastny, K. Ye. Samouylov

*Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

The Poisson point process (PPP) is widely used in performance analysis of wireless communications technologies as a basic model for random deployment of communicating entities. The reason behind widespread use of PPP is analytical tractability in terms of closed-form distributions of distances to the n -th neighbour needed for performance analysis. At the same time, the process allows for infinitesimally close distances between communicating stations not only contradicting the reality but presenting fundamental difficulties in analysis when

used with power-law propagation models. As an alternative suggested in the literature and free of abovementioned deficiencies are the hardcore processes where a certain separation distance between points is always presumed. Unfortunately, no closed-form expressions for distance distributions is available for these processes. We study distance distributions of Matern hardcore process and propose analytical approximations based on acyclic phase type distributions. The nature of approximation as a mixture of exponentials allows for their use in analytical performance analysis. Results for a range of process intensities are reported.

Keywords: hardcore process, approximation, Poisson process, distance distribution, SINR.

УДК 004.057.4

Исследование зависимости пропускной способности беспроводного канала связи от параметров настройки точки доступа

Л. И. Абросимов, М. А. Руденкова

Национальный исследовательский институт «МЭИ»

Аннотация. В докладе приводится методика измерения с помощью протокола SNMP пропускной способности беспроводного канала связи в составе корпоративной вычислительной сети (КВС) для оценки зависимости от варьируемых параметров настройки точки доступа при синхронизации измерений относительно системного времени точки доступа. Приводятся результаты измерений для трафика организованного протоколами TCP и UDP. Результаты измерений используются для адаптации пропускной способности сегментов беспроводной сети к требованиям передаваемого трафика КВС.

Ключевые слова: корпоративные беспроводные сети, оценка пропускной способности.

1. Введение

Развитие беспроводных технологий стандарта IEEE 802.11 приводит к увеличению разнородных пользовательских устройств IEEE 802.11 внутри корпоративной беспроводной ЛВС (КБЛВС). В БКЛВС требуется обеспечить согласованность предоставляемой пропускной способности и трафика пользовательских приложений. Для этого требуется осуществлять измерения характеристик передаваемых данных и их обработки сетевыми устройствами. КБЛВС представлена базовыми наборами обслуживания (от англ. - Basic Service Set, BSS) и коммутационной сетью, обеспечивающей доступ к ресурсам корпоративной сети или выход в сеть Интернет. Требуются средства для оценки пропускной способности, обеспечиваемой каждым BSS. Для оценки пропускной способности используются захват и анализ трафика или получение статистики системных счетчиков входных/выходных кадров/пакетов с сетевого устройства. В данной работе для определения пропускной способности была разработана методика на основе протокола простого сетевого управления (от англ. - Simple Network Management Protocol, SNMP) для синхронизации полученных данных статистики от сетевого устройства, учитывалось значение системного времени сетевого устройства. Также с помощью разработанной методики производилось варьирование параметров канального уровня для выявления зависимости пропускной способности от данных параметров для конкретного BSS. Измерения значения пропускной способности от значений конкретного параметра канального уровня позволяет производить адаптивную настройку BSS.

2. Методика измерения пропускной способности

2.1. Формулировка задачи

Рассматривается беспроводная сеть стандарта 802.11 работающая в инфраструктурном режиме (Центральное устройство данной сети точка доступа AP (от англ. - Access Point) образует с подключенными станциями STA (от англ. - Station) беспроводной канал общего пользования, передача данных происходит только через AP). AP устанавливает значения параметров $\{z_k^*\}$ данного канала $\{p_k\}$ ($k = \overline{1, n}$) и управляет передачей параметров из беспроводной сети в проводную. Таким образом, характеристики беспроводного канала передачи данных $\{r_i\}$ ($i = \overline{1, l}$) определяются беспроводным интерфейсом AP, а характеристики переданных/принятых данных в проводную среду $\{e_j\}$ ($j = \overline{1, r}$) проводным интерфейсом AP. Требуется определить набор зависимостей $r_{ki}(z_k)$ и $e_{kj}(z_k)$.

2.2. Этапы методики измерения характеристик сетевых интерфейсов

1. Определить набор характеристик интерфейсов AP:
 - (а) для беспроводного интерфейса стандарта IEEE802.11 – $\{r_i\}$;
 - (б) для проводного интерфейса стандарта IEEE802.3 – $\{e_j\}$.
2. Синхронизация времени Server и AP. Зафиксировать системное локальное время Server T_s^0 и отправить запрос SNMP GET на AP для установки системного времени T_{AP}^0 ;
3. Сформировать запрос SNMP BULK для набора $\{r_i\}$;
4. Зафиксировать системное локальное время Server T_s^1 и отправить сформированный в п.3. методики запрос SNMP BULK. Зафиксировать полученные значения $\{x_i\}$ для набора $\{r_i\}$. $T_{AP}^1 = T_{AP}^0 + (T_s^1 - T_s^0) [ms]$
5. Сформировать запрос SNMP BULK для набора $\{e_i\}$.
6. Зафиксировать системное локальное время Server TS2 и отправить сформированный в п.5. методики запрос SNMP BULK [1]. Зафиксировать полученные значения $\{y_j\}$ для набора $T_{AP}^2 = T_{AP}^1 + (T_s^2 - T_s^1) [ms]$.

2.3. Этапы методики изменения параметров канального уровня

1. Определить набор регулируемых параметров канального уровня для исследуемой BSS $\{p_k\}$.
2. Определение набора значений $\{z_k^*\}$ для параметров $\{p_k\}$ ($k = \overline{1, n}$);

3. Определить значение набора значений $\{z_k\}$ регулируемых параметров $\{p_k\}$;
4. Используя функцию SNMP SET установить определенные значения $\{z_k\}$ регулируемых параметров $\{p_k\}$ [1].

2.4. Этапы методики измерения пропускной способности

1. Используя методики п. 2.2. и п. 2.3. определить наборы характеристик $\{r_i\}$ и $\{e_j\}$, регулируемых параметров $\{p_k\}$ и их набор значений $\{z_k^*\}$;
2. На шаге s определить значения набора параметров $\{z_k^s\}$;
3. Используя методику п. 2. 3. и методику п. 2. 2, установить значения параметров $\{z_k^s\}$ и получить значения характеристик интерфейсов $\{r_i^s\}$ и $\{e_j^s\}$ как функцию $f(z_k^s)$, последовательно получить функцию значения характеристик интерфейсов $\{r_i^s\}$ и $\{e_j^s\}$ последовательно изменяя значения набора $\{z_{k+m}^s\}$ параметра $p_{k+m}^s (m = \overline{1, n})$. Пропускную способность BSS определим из характеристик проводного интерфейса, с помощью которого трафик пользователей передается по восходящем/нисходящим каналам КЛВС:

$$C_{in} = \frac{\Delta e_{in}}{\Delta t}; C_{out} = \frac{\Delta e_{out}}{\Delta t}; C = C_{in} + C_{out};$$

где C - пропускная способность интерфейса, C_{in} - для входящего трафика, C_{out} - для исходящего трафика; e_{in} - количество входящих октетов трафика на проводном интерфейсе AP; e_{out} - количество исходящих октетов трафика на проводном интерфейсе AP;

3. Реализация методики измерения пропускной способности

Рассматривается тестовый стенд беспроводной корпоративной сети стандарта IEEE 802.11 для измерения пропускной способности рис. 1.:

1. беспроводная точка доступа AP, представленная Cisco AIR-LAP1232AG; Интерфейсы AP:
dot11radio0 – интерфейс беспроводной сети стандарта IEEE 802.11;
fa0/0 – интерфейс проводной сети стандарта IEEE 802.3.
2. мобильный компьютер STA, представленный компьютером с беспроводным адаптером Cisco Aironet 802.11a/b/g Wireless Adapter;
3. сервер Server, представленный компьютером, подключенным с помощью интерфейса стандарта IEEE 802.3 к КЛВС;

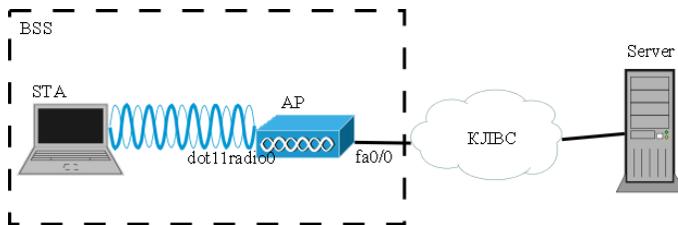


Рис. 1. Схема стенда измерения пропускной способности

Требуется: С помощью разработанных методик определить зависимость пропускной способности от значений регулируемых параметров канального уровня.

Этап. 1. Набор характеристик интерфейсов AP, получаемых по протоколу SNMP:

- а. для беспроводного интерфейса стандарта IEEE802.11, стандартные параметры из IEEE802dot11-MIB [2];

dot11TransmittedFragmentCount – счетчик успешно переданных MPDU с индивидуальным адресом или с групповым адресом с данными или управляющие;

dot11MulticastTransmittedFrameCount – счетчик успешно переданных MSDU с групповым адресом;

dot11FailedCount – счетчик увеличивается при ошибке передачи MSDU по истечении таймера дополнительных попыток, определяемого параметром dot11ShortRetryLimit или dot11LongRetryLimit;

dot11RetryCount – счетчик увеличивающийся при успешной передаче MSDU после одной или нескольких попыток повторной передачи;

dot11MultipleRetryCount – счетчик, увеличивающийся при успешной передаче MSDU после нескольких попыток повторной передачи;

dot11FrameDuplicateCount – счетчик дублированных фреймов;

dot11RTSSuccessCount – счетчик ответов CTS на запросы RTS;

dot11RTSFailureCount – счетчик неполученных CTS на запросы RTS;

dot11ACKFailureCount – счетчик неполученных ожидаемых ACK;

dot11ReceivedFragmentCount – счетчик успешно принятых MPDU с индивидуальным адресом или с групповым адресом с данными или управляющие;

dot11MulticastReceivedFrameCount – счетчик успешно принятых MSDU с групповым адресом;

dot11FCSErrorCount – счетчик MPDU с ошибочной контрольной последовательностью;

dot11TransmittedFrameCount – счетчик всех успешно переданных MSDU.

b. – интерфейс проводной сети стандарта IEEE 802.3. [2]

ifInOctets – счетчик количества входящих октетов трафика;

ifInUcastPkts – счетчик количества входящих одноадресных пакетов обработанных на данном подуровня для передачи на вышестоящие (под-) уровни;

ifInNUcastPkts – счетчик количества входящих пакетов групповой или широковещательной рассылки обработанных на данном подуровня для передачи на вышестоящие (под-) уровни;

ifInDiscards – счетчик количества входящих отброшенных пакетов, даже если не было обнаружено никаких ошибок. Одной из возможных причин может быть освобождение буфера;

ifInErrors – счетчик количества входящих отброшенных пакетов, содержащих ошибки;

ifOutOctets – счетчик количества выходящих октетов трафика;

ifOutUcastPkts – счетчик количества выходящих одноадресных пакетов обработанных на данном подуровня для передачи на вышестоящие (под-) уровни;

ifOutNUcastPkts – счетчик количества выходящих пакетов групповой или широковещательной рассылки обработанных на данном подуровня для передачи на вышестоящие (под-) уровни;

ifOutDiscards – счетчик количества выходящих пакетов групповой или широковещательной рассылки обработанных на данном подуровня для передачи на вышестоящие (под-) уровни;

ifOutErrors – счетчик количества выходящих отброшенных пакетов, содержащих ошибки.

Набор регулируемых параметров беспроводного интерфейса AP, управляемых с помощью протокола SNMP : [2]

dot11CFPPeriod – Устанавливает через какое количество интервалов DTIM начнется CFP (contention-free period – период работы беспроводной среды без механизма предотвращения коллизий CSMA/CA);

dot11CFPMaxDuration – Максимальная продолжительность режима CFP;

dot11PowerManagementMode – Определяет возможность в режиме экономии энергии. Данный режим влияет на производительность, т. к. включая его клиенты могут буферизировать свои кадры вместо активного доступа к беспроводного канала, освобождая доступ другим станциям;

- dot11OperationalRateSet** – Определяет набор скоростей для подключаемых станций;
- dot11BeaconPeriod** – Устанавливает период времени посылки кадров Beacon;
- dot11DTIMPeriod** – Период через который будет передан кадр Beacon с DTIM Count = 0;
- dot11RTSThreshold** – Количество октетов в MPDU (medium access control (MAC) protocol data unit (фрагмент данных протокола уровня доступа к среде)) для которого не требуется механизм RTS/CTS;
- dot11ShortRetryLimit** – Максимальное количество попыток передать кадр, размер кот. меньше или равен dot11RTSThreshold, прежде чем определить невозможность передачи;
- dot11LongRetryLimit** – Максимальное количество попыток передать кадр, размер кот. Больше dot11RTSThreshold, прежде чем определить невозможность передачи;
- dot11FragmentationThreshold** – Максимальное количество октетов MPDU, которые будут переданы на физический уровень;
- dot11MaxTransmitMSDULifetime** – Максимальное время в TU, после которого дальнейшие попытки передать кадр будут прерваны;
- dot11MaxReceiveLifetime** – Максимальное время в TU, после которого дальнейшие попытки принять кадр будут прерваны.

Этап 2. Начальные значения регулируемых параметров канального уровня в начале методики (см. табл. 2). Этап 3. Шаг 1. На Server для данной методики написана программа на языке C, использующая библиотеку NET-SNMP для отправки и получения запросов по протоколу SNMP Используя методику п. 2.3. последовательно устанавливаем значения параметров табл. 2. Для синхронизации времени AP и Server согласно методике п. 2. 2. Используется функция `gettimeofday()` для Server (мкс), при получении сообщения SNMP GET системного времени AP – `sysUptime` - Время от начала работы ТД в тиках и в сотых долях миллисекунд после последней перезагрузки (SNMPv2-MIB:`sysUpTimeInstance`). По методике п. 2. 2. и 2.3. Последовательно изменяем значение параметра $p1$ и фиксируем значение характеристик интерфейсов и производим расчет пропускной способности $f(z_1^1)$. Переходим к параметру $p2$, устанавливаем следующее доступное значение параметру $p2$. По методике п. 2. 2. и 2.3. Последовательно изменяем значение параметра $p1$ и фиксируем значение характеристик интерфейсов и производим расчет пропускной способности $f(z_1^2)$ и так далее.

Таблица 1

Регулируемые параметры канального уровня и их диапазон значений

Об-ние	Название	Значение
p1	dot11CFPPeriod	[0-255]/количество DTIM интервалов
p2	dot11CFPMaxDuration	[0-65535]/количество TU
p3	dot11PowerManagementMode	[active(1),powersafe(0)]
p4	dot11OperationalRateSet	[1-126]/bytes
p5	dot11BeaconPeriod	[1-65535]/количество TU
p6	dot11DTIMPeriod	[1-255]/количество dot11BeaconPeriod
p7	dot11RTSThreshold	[0-2347]/количество октетов
p8	dot11ShortRetryLimit	[1-255]/количество попыток
p9	dot11LongRetryLimit	[1-255]/количество попыток
p10	dot11FragmentationThreshold	[256-2346]/количество октетов
p11	dot11MaxTransmitMSDULifetime	[1-4294967295]/ количество TU
p12	dot11MaxReceiveLifetime	[1-4294967295]/ количество TU

Таблица 2

Регулируемые параметры канального уровня и их значения

Параметр	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12
Значение	0	0	0	1	1	1	0	1	1	256	1	1

4. Результаты

Для проверки методики был собран тестовый стенд см. рис. 1. На STA запущен iperf client для генерации трафика TCP и UDP с характеристиками реальной беспроводной сети из работы [3]. В докладе приводятся графики полученных зависимостей.

5. Заключение

Разработанная методика позволяет точно фиксировать значение характеристик, используя синхронизацию локального системного времени Server и системного времени ТД. Используя запрос GETBULK – увеличивается точность значений, т.к. значения передаются по сети в одном пакете и одним запросом. Данную методику можно использовать в реальной корпоративной сети для определения состояния беспроводного канала, устанавливая в такой сети программно-аппаратные средства анализа используемых протоколов можно использовать данную методику для проведения адаптивной настройки ТД. Также в данной методике использовались стандартные базы параметров (MIB) для параметров канального уровня, характеристик интерфейса Ethernet ТД. Использование стандартных баз позволяет использовать данную методику с любыми устройствами разных производителей поддерживающих протокол SNMP, в настоящее время данный протокол обязателен для устройств корпоративного уровня, что также оправдывает выбор данной методики для корпоративных БЛВС.

Литература

1. *Net-SNMP Project Manual pages for SNMP.* //Net-SNMP Project, 2016. — <http://www.net-snmp.org/docs/man/>.
2. *Pierrick S. SNMP MIB Resource* //SNMPlink.org. — <http://www.snmplink.org/resource/mib/>.
3. *Na C., Chen J. K., Rappaport T. S.* Measured Traffic Statistics and Throughput of IEEE 802.11b Public WLAN Hotspots with Three Different Applications. — *Wireless Communications, IEEE Transactions on In Wireless Communications, IEEE Transactions on*, Vol. 5, No. 11. (19 December 2006). — P. 3296–3305.

UDC 004.057.4

Analysis of Throughput Wireless Media and Settings for Access Point Data Layer

L. I. Abrosimov, M. A. Rudenkova

National research university "MPEI", Moscow, Russia

In this paper describes methodology for measuring throughput of wireless LAN, using SNMP protocol. Also this methodology describes how sets settings for access point and studies how this modification can change throughput of wireless LAN.

Keywords: corporate wireless networks, throughput analysis .

УДК 519.691

Минимизация затрат при создании системы мониторинга дата-центра при детерминированном потоке данных

Л. И. Абросимов, А. А. Ларин

*Национальный исследовательский институт "МЭИ", Москва, Россия,
National Research University "Moscow Power Engineering Institute Moscow,
Russia,*

Аннотация. Рассматривается система мониторинга, циклически собирающая данные о состоянии технических средств датацентра с известным периодом опроса. Разработаны математические соотношения, учитывающих особенности функционирования датацентра и системы мониторинга, на основании которых предложена методика оптимизации решения задачи минимизации затрат при построении системы мониторинга состояния технических средств. В докладе приводятся примеры использования методики.

Ключевые слова: беспроводная сеть дата-центра, система мониторинга дата-центра.

1. Введение

С каждым годом растут потребности бизнеса в дата-центрах (ДЦ), соответственно, необходим всё более качественный уровень предоставления услуг и вместе с этим усложняется сетевая инфраструктура центров обработки данных. Владельцам ДЦ необходимо бесперебойно предоставлять всю необходимую инфраструктуру, для этого необходим постоянный контроль систем энергоснабжения, пожаротушения, кондиционирования, а также слежение за аппаратно-техническими средствами.

При разработке и модернизации ДЦ возникают следующие проблемы:

- Необходимо построение эффективной автоматизированной или автоматической системы управления процессами в центрах обработки данных;

- Должны быть построены системы мониторинга, которые обеспечат необходимое время реакции в случае внештатной ситуации.

Для этого необходимо присутствие специальных вычислительных сетей, которые смогут обеспечить необходимые показатели. Также в случае автоматизированных систем, необходимо удобное для операторов отображение текущего состояния технических средств.

В настоящее время, с развитием беспроводных технологий, есть возможность оперативного сбора различной информации, при этом время на построение такой сети уменьшается по сравнению с проводными реализациями, когда необходимо иметь определенный запас проводной среды или же необходимо время на прокладку новых проводов.

Обычно для анализа состояния сети используется опрос устройств по протоколу SNMP, а также косвенный анализ, исходя из обмена управляющей информацией между сетевыми узлами, например, в ситуации, когда теряется канал между двумя устройствами.

На рисунке 1 представлен машинный зал ДЦ с мониторингом на основе проводной среды.

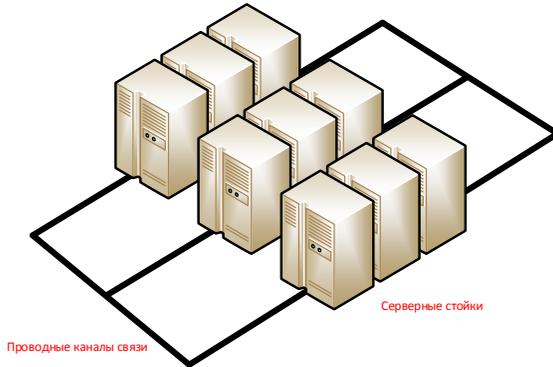


Рис. 1. Машинный зал с проводной средой мониторинга

В этом случае, возможна ситуация, когда в результате обрыва проводов теряются как данные, так и информация о состоянии устройств, что усложняет определение аварийной ситуации.

Беспроводная же технология позволяет осуществлять опрос независимо от основной сетевой инфраструктуры и, следовательно, образуется таким образом отдельный сетевой слой мониторинга.

На рисунке 2 представлен машинный зал ДЦ с мониторингом на основе беспроводных датчиков.

Так современные производители предлагают в сетевом и серверном оборудовании иметь установленные производителями датчики, с которых будет возможность, после обработки, выводить информацию на экран операторам.

Циклически опрашивая датчики, можно точно узнавать и обрабатывать необходимые данные, а тот датчик, который не отвечает на очередном этапе сбора есть возможность оперативно локализовать.

2. Минимизация затрат

В ДЦ для клиента составляется соглашение об уровне предоставления услуги (SLA), например, доступность сетевого сервиса или доступность серверов. И определяются такие параметры как:

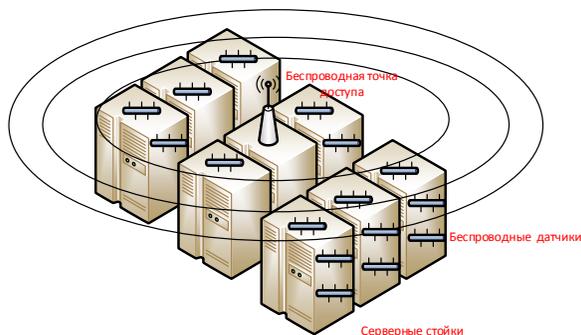


Рис. 2. Машинный зал с беспроводным мониторингом

Согласованное время работоспособности (СВР) – отражает временной период, в рамках которого сервис должен функционировать в нормальном режиме. Например: круглосуточная доступность, 7 дней в неделю.

Согласованное время поддержки (СВП) – указывает на временной период, в который осуществляется поддержка сервиса со стороны ДЦ. Например: 10:00-18:00, понедельник-пятница.

Период оказания услуги (ПОУ) – период, в который сервис будет предоставляться. Например, календарный месяц.

Время реакции (ВР) – определяется как время от момента получения информации о функциональном отказе (от системы мониторинга или от клиента) до момента старта восстановительных работ по восстановлению работоспособности сервиса.

Время простоя (ВП) – суммарное время простоя в работе сервиса за период оказания услуги.

Доступность (Д) – максимальный процент доступности сервисов за период оказания услуги.

Таким образом доступность определяется как:

$$Д = \frac{T_{\text{СВР}} - T_{\text{ВП}}}{T_{\text{СВР}}} \cdot 100\%$$

Например, при суммарном ВП равному 3 часа, при ПОУ - 30 дней, СВР - 24 часа, 7 дней в неделю:

$$Д = \frac{30 \cdot 24 \cdot 60 - 3 \cdot 60}{30 \cdot 24 \cdot 60} \cdot 100\%$$

$$Д = 99.58\%$$

Доступность за период указывает на нарушение работоспособности и доступности сервисов и, следовательно, определяет размер денежных компенсаций клиенту. В таблице 1 приведен пример процента компенсации (относительно стоимости предоставления сервиса) при различных показателях доступности таких сервисов, как предоставленные сети и серверной инфраструктуры.

Таблица 1
Пример показателей доступности к размеру компенсации

Показатели доступности (%)	Размер компенсации (%)
99.91 > Д > 9.73	5
99.73 > Д > 99.44	10
99.44 > Д > 98.9	15
98.9 > Д	20

Затраты ДЦ (Здц) напрямую зависят от размера компенсаций.

Таким образом, необходимо увеличивать показатель доступности. При этом доступность зависит от времени реакции:

$$Д \sim Твр.$$

А время реакции зависит от времени опроса соответствующих технических средств:

$$Твр \sim Топ.$$

При помощи автоматизированной системы мониторинга на основе беспроводной сети можно увеличить доступность по сравнению с традиционным проводным мониторингом, уменьшив тем самым возможные убытки для владельцев ДЦ.

На рисунке 3 представлена беспроводная сеть ДЦ, в которой датчики опрашиваются циклически (детерминированный поток данных).

Происходит опрос каждого датчика за время Td_i , опрашивающий узел получает Qd_i байт данных. Полученные данные передаются через цепочку промежуточных узлов аккумулирующему серверу. Между каждым датчиком и узлом – беспроводной канал связи (БКС) с шириной полосы пропускания W . В сети Nd датчиков.

Учитывая цикличность опрос датчика происходит с квантом Tq на каждый датчик.

Полный цикл получения данных состоит из цикла опроса датчиков, передачи по проводным каналам связи (ПКС), обработке на аккумулярующем сервере (С):

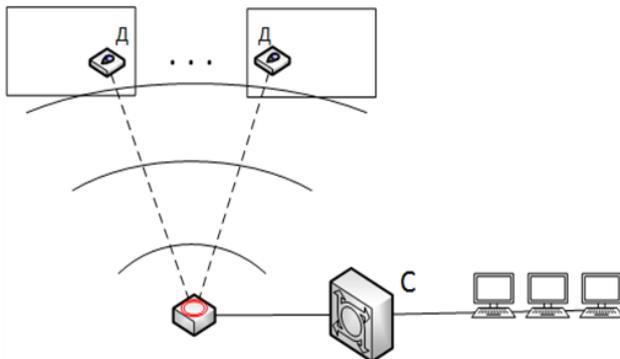


Рис. 3. Беспроводная система мониторинга ДЦ

$$T_{\text{ц}} = T_{\text{ц,опроса}} + T_{\text{кс}} + T_{\text{обр.с}}$$

Общее время цикла опроса датчиков:

$$T_{\text{ц,опроса}} = Nd \cdot Tq$$

При этом должно выполняться условие:

$$Nd \cdot Tq > \sum_{i=1}^{Nd} \left(\frac{Qd_i + Qs}{W_{\text{бкс}}} + Td_i \right)$$

Время передачи по проводным каналам связи серверу:

$$T_{\text{кс}} = \frac{\sum_{i=1}^{Nd} (Qd_i + Qs)}{W_{\text{пкс}}}$$

Время обработки на аккумулярующем данные с датчиков сервере зависит от производительности сервера:

$$T_{\text{обр.с}} = \frac{\sum_{i=1}^{Nd} Qd_i}{Pc}$$

Таким образом, полный цикл:

$$T_{\text{ц}} = Nd \cdot Td + \frac{\sum_{i=1}^{Nd} (Qd_i + Qs)}{W_{\text{пкс}}} + \frac{\sum_{i=1}^{Nd} Qd_i}{Pc}$$

Время реакции на событие складывается из реакции оператора (при автоматизированной системе) и времени получения данных (через период цикла):

$$T_{вр} = T_{ц} + T_{реак}$$

А время простоя определяемое в SLA определяется как сумма времени реакции и времени восстановления:

$$T_{вп} = T_{вр} + T_{восст}$$

Таким образом, при переходе к беспроводному опросу датчиков, уменьшается $T_{ц}$, а следовательно, это уменьшает время простоя.

Учитывая параметры (Qd_i , Td_i , $W_{бкс}$, $W_{пкс}$, Nd , Pc), уменьшение $T_{ц}$ минимизирует затраты ДЦ:

$$\text{при } T_{ц} \downarrow, Z_{дц} \rightarrow \min$$

3. Заключение

Благодаря детерминированному потоку есть возможность решить несколько задач: контролировать каждый датчик, а в случае отсутствия данных оперативно реагировать через автоматизированную систему. Переход от мониторинга по существующим каналам связи к беспроводному мониторингу датчиков в оборудовании улучшает показатели доступности сервиса и, тем самым, минимизирует затраты ДЦ.

В дальнейшем возможен переход к автоматической системе, которая позволила бы дополнительно уменьшить время восстановления за счёт автоматизации процессов, например, за счёт ввода резервов при авариях.

Литература

1. Zhang L., Ardagna D. SLA Based Profit Optimization in Web Systems — 13th International Conference on World Wide Web (WWW'04) , New York, USA, 2004.
2. Столлингс В. Беспроводные линии связи и сети.: Пер. с англ. — М.: Издательский дом Вильямс, 2003. - 640 с.

UDC 519.691

Minimization of Data Center Cost Creating Monitoring System with Determinate Data Flow

L. I. Abrosimov, A. A. Larin

The article consider monitoring system with cyclic poll of data center equipment status data with definite polling period. There are developed mathematical relationships in view data center and monitoring system functional features. Optimisation methods of data center cost with monitoring system based on this relationships was offered. Examples of using methods consider in report.

Keywords: wireless network of data center, monitoring system of data center.

УДК 517.9

О задаче об успокоении системы управления с последствием

А. Ш. Адхамова, А. Л. Скубаческий

*Кафедра прикладной математики,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

Аннотация. Рассматривается задача об успокоении управляемой системы с последствием, описываемой системой дифференциально-разностных уравнений нейтрального типа. Исследуется связь между вариационной задачей для нелокальных функционалов, описывающих многомерную систему управления с запаздыванием, и соответствующей краевой задачей для системы дифференциально-разностных уравнений, исследуется вопрос о существовании и единственности обобщенного решения данной краевой задачи.

Ключевые слова: система управления с последствием, нелокальные краевые задачи, обобщенное решение.

1. Введение

В последние годы интерес к дифференциально-разностным уравнениям связан с приложениями в таких областях, как теория многослойных пластин и оболочек в авиации, теории многомерных диффузионных процессов в теории плазмы, теории нелинейных лазерных систем и других. В работе [1] Н. Н. Красовский сформулировал и изучил задачу об успокоении системы с последствием, описываемой дифференциально-разностным уравнением запаздывающего типа. Он свел эту задачу к краевой задаче для системы дифференциально-разностных уравнений с отклоняющимся аргументом в младших членах. Краевые задачи для дифференциально-разностных уравнений со сдвигами аргумента в старших членах были изучены в работах [2, 3]. А. А. Скубаческий в работе [4] обобщил задачу Н. Н. Красовского на случай, когда уравнение, описывающее управляемую систему, содержит также старшие члены с запаздыванием. Но во всех выше указанных работах рассматривалась одномерная задача. Реальной системе управления соответствует многомерная модель. В работе [5] Д. Д. Леоновым рассматривается обобщение задачи, изученной А. Л. Скубаческим, когда число входов системы управления равно числу выходов, но при этом коэффициент при старшем члене без запаздывания уравнения равен единице. В этой работе рассматривается линейная система управления, описываемая системой дифференциальных уравнений с запаздыванием, где коэффициенты – квадратные постоянные n -мерные матрицы.

2. Основная часть

Рассмотрим линейную систему управления, описываемую системой дифференциальных уравнений с запаздыванием

$$Ay'(t) + By'(t - \tau) + Cy(t) + Dy(t - \tau) = u(t), t > 0.$$

$$y(t) = \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix}, u(t) = \begin{pmatrix} u_1(t) \\ \vdots \\ u_n(t) \end{pmatrix},$$

где A, B, C, D — квадратные n -мерные постоянные матрицы, запаздывание $\tau > 0$ — константа, $u(t)$ — управление. Предыстория этой системы определяется начальным условием

$$y(t) = \varphi(t), t \in [-\tau, 0], \quad \varphi(t) = \begin{pmatrix} \varphi_1(t) \\ \vdots \\ \varphi_n(t) \end{pmatrix}, \quad (1)$$

где $\varphi(t)$ — заданная функция.

Мы рассмотрим задачу о приведении системы (1), (2) в положение равновесия. Найдем такое управление $u(t)$, что:

$$y(t) = 0, t \in [T - \tau, T], \quad (2)$$

где $T > 2\tau$.

Вектор-функция $y(t)$, удовлетворяющая условиям (1) — (2), не единственна. Поэтому также будем предполагать, что

$$\int_0^T \|u(t)\|^2 dt \rightarrow \min,$$

где $\|\cdot\|$ — евклидова норма. Таким образом, мы получаем вариационную задачу о минимуме функционала энергии

$$J(y) = \int_0^T \|Ay'(t) + By'(t - \tau) + Cy(t) + Dy(t - \tau)\|^2 dt. \rightarrow \min \quad (3)$$

с краевыми условиями (1), (2).

Пусть $y \in W^{1,n}(-\tau, T)$ — решение вариационной задачи (1) — (3), где $\varphi \in W^{1,n}(-\tau, 0)$. Обозначим

$$\widetilde{W} = \{v \in W^{1,n}(-\tau, T) : v(t) = 0, t \in [-\tau, 0] \cup [T - \tau, T]\}.$$

Пусть $v(t) \in \widetilde{W}$ — произвольная фиксированная функция.

Определение 1. Функцию $y \in W^{1,n}(-\tau, T)$ назовем обобщенным решением краевой задачи

$$\begin{aligned} & -[y'^T(A^T A + B^T B) + y'^T(t - \tau)B^T A + y'^T(t + \tau)A^T B]' + \quad (4) \\ & + y'^T(t - \tau)B^T C + y'^T(t + \tau)A^T D + y'^T(t)(B^T D + A^T C) + \\ & + y^T(t)(C^T C + D^T D) + y^T(t - \tau)D^T C + y^T(t + \tau)C^T D - y'^T(t)(C^T A + \\ & + D^T B) - y'^T(t - \tau)D^T A - y'^T(t + \tau)C^T B = 0, t \in (\tau, T - \tau) \end{aligned}$$

с краевыми условиями (1), (2), если выполняется условие

$$y'^T(A^T A + B^T B) + y'(t - \tau)B^T A + y'^T(t + \tau)A^T B \in W^{1,n}(\tau, T - \tau)$$

и функция $y(t)$ удовлетворяет системе уравнений (4) и краевым условиям (1), (2).

Утверждение. Функция $y \in W^{1,n}(-\tau, T)$ является обобщенным решением задачи (4), (1), (2) тогда и только тогда, когда она удовлетворяет интегральному тождеству

$$\begin{aligned} B(y, v) = & \int_0^T \{ [y'^T(A^T A + B^T B) + y'^T(t - \tau)B^T A + \\ & + y'^T(t + \tau)A^T B]v'(t) + [y'^T(t - \tau)B^T C + y'^T(t + \tau)A^T D + \\ & + y'^T(t)(B^T D + A^T C) + y^T(t)(C^T C + D^T D) + y^T(t - \tau)D^T C + \\ & + y^T(t + \tau)C^T D - y'^T(t)(C^T A + D^T B) - y'(t - \tau)D^T A - \\ & - y'^T(t + \tau)C^T B]v(t) \} dt = 0 \end{aligned}$$

для всех $v \in \delta W^{1,n}(0, T - \tau)$ и краевым условиям (1), (2).

Были получены следующие результаты:

Теорема 1. Функция $y \in W^{1,n}(-\tau, T)$ реализует минимум функционала (3) с краевыми условиями (1), (2) тогда и только тогда, когда она является обобщенным решением краевой задачи (4), (1), (2).

Теорема 2. Для любого $\varphi \in W^{1,n}(-\tau, 0)$ существует единственное обобщенное решение $y \in W^{1,n}(-\tau, T)$ краевой задачи (4), (1), (2), при этом

$$\|y\|_{W^{1,n}(-\tau, T)} \leq c \|\varphi\|_{W^{1,n}(-\tau, 0)},$$

где $c > 0$ и не зависит от φ .

Литература

1. *Красовский Н. Н.* Теория управления движением. — М.:Наука, 1968.
2. *Каменский Г. А., Мышкис А. Д.* К постановке краевых задач для дифференциальных уравнений с отклоняющимся аргументом и несколькими старшими членами. // Дифференц. уравн. — 1974. — Т. 10, №3. — С. 409-418.
3. *Каменский Г. А.* Краевые задачи для уравнений с формально симметричными дифференциально-разностными операторами. // Дифференц. Уравн. — 1976. — Т. 12, № 5. — С. 815-824.
4. *Skubachevskii A. L.* Elliptic functional differential equations and applications. // -Basel-Boston-Berlin:Birkhauser. — 1997.
5. *Леонов Д. Д.* К задаче об успокоении системы управления с последствием // Современная математика. Фундаментальные направления. — 2010. — Том 37. — С. 28–37.

UDC 517.9

About damping problem for control system with delay

A. S. Adkhamova, A. L. Skubachevskii

*Department of Applied Mathematics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

We consider damping problem for control system with delay described by the system of differential-difference equations of neutral type, complete the proof of the theorem on the relationship of the variational problem for the nonlocal functionals and the corresponding boundary value problem, prove the existence and uniqueness of generalized solutions of the boundary value problem for the corresponding system of functional differential equations.

Keywords: control system with delay, nonlocal boundary value problem, generalized solution.

UDC 004.414.23

Characteristics of Priority Queues with High Utilization Parameter

T. I. Aliev*, V. V. Sosnin*[†]

* *Department of Computer Science
ITMO University*

Kronverksky Pr. 49, St. Petersburg, 197101, Russia

[†] *Software Development Department
Research Institute "Mashtab"*

Kantemirovskaya 5A, St. Petersburg, 194100, Russia

Abstract. Queuing systems with priority-based queuing disciplines are considered for heavy traffic conditions when utilization ratio is near 1. We analyze properties of systems represented by models with infinite buffer capacity using simulation and analytical modeling. We offer a method to approximately estimate metrics of such queuing systems using two moments of probability distributions of respective system parameters. To calculate sojourn time variation (jitter) and estimate buffer capacity for a given loss probability, the multiexponential probability distributions are used for approximation.

Keywords: queuing theory, priority queues, infinite capacity queuing nodes, utilization ratio, heavy traffic queues, queuing discipline, sojourn time, jitter.

1. Introduction

Analytical priority queuing models are typically used to analyze computer networks with multiservice kinds of traffic that have varying Quality of Service (QoS) requirements [1–3]. A request in such models usually represents a protocol data unit (e.g. an Ethernet frame or an IP packet) traversing a switch or a router which thus represents the server of the queuing model. QoS requirements for each kind of traffic usually differ and are typically represented as constraints on the following metrics: probability of packet loss due to limited queuing buffer capacity, average sojourn time and sojourn time jitter (delay variation in terms of [4]). To satisfy these constraints individually for each class, different priority disciplines are used, along with specific buffer space allocation policies.

Most of current results in the field (see [5, 6]) are analytical models for one-processor stationary multi-class queues with offered load below 1 with no means to calculate delay variation. At the same time, processes in computer networks often impose non-stationary high loads on network devices. In this paper we consider a queuing system to be highly loaded if utilization parameter exceeds 0.9.

Simulation of highly loaded queuing systems has known convergence issues. This can be illustrated using M/M/1 model in which an insignificant increase of utilization parameter from 0,98 to 0,99 leads to an astonishing two-fold increase of average sojourn time. It means that simulation results

will have a very wide confidence interval. So the preferred method is analytical modeling that allows getting either exact results or upper/lower bounds for QoS metrics.

In real systems with high utilization, requests may be rejected due to a limited physical buffer capacity thus leading to a low system efficiency. Using infinite buffer capacity queuing models (IBCQM) to analyze such systems is impossible because performance metrics in these models are undefined for utilization parameter close to 1. On the other hand, using finite buffer capacity queuing models (FBCQM) is also problematic because these models have a very limited scope: they exist only for exponentially distributed inter-arrival and service times with no priority segregation. Our idea is to use FBCQM to estimate buffer capacity of IBCQM for a given upper bound of loss probability.

In this paper, multiclass single server priority queues with infinite buffer capacity are considered for the so-called mixed-priority queuing discipline. This discipline allows mixing non-preemptive priority (NPP) and FCFS queuing rules within a single queuing system. It enables aggregating several request classes into FCFS groups between which non-preemptive priority queuing may be applied. We analyze how performance metrics of different classes depend on the utilization parameter. The metrics being analyzed are average sojourn time, sojourn time jitter (delay variation) and loss probability.

Examples of real-life systems that could be modeled by described queuing systems are a network router and a data link in a computer network. In the latter case, the model describes the process during which link layer frames of different classes are waiting in a NIC buffer before being sent. Since the QoS requirements of respective classes may to be different, one has to use priority queuing in order to differentially satisfy individual requirements.

2. Statement of the Problem

Consider a queuing system with H classes of requests each arriving according to Poisson process with rates $\lambda_1, \dots, \lambda_H$ per unit time respectively. Service time τ_{b_h} for requests of h -th class ($h = 1, \dots, H$) has a general distribution for which at least three moments are known: $b_h = b_h^{(1)}, b_h^{(2)}$ and $b_h^{(3)} : b_h^{(s)} = \int_0^\infty \tau_{b_h}^s dF(\tau_{b_h})$, where $F(\tau_{b_h})$ is the cumulative distribution function (CDF) of τ_{b_h} . Requests are serviced in the order defined by a mixed-priorities matrix (see [7]): $Q = [q_{ij}]$ ($i, j = 1, \dots, H$), where $q_{ij} = 1$ if i -th class has non-preemptive priority over j -th class and $q_{ij} = 0$ if i -th and j -th class form an FCFS (First-Come-First-Served) group. Matrix Q is considered consistent if

- elements in every row form a non-decreasing sequence as row index increases;

— elements in every column form a non-increasing sequence as column index increases.

It means that higher class index always means lesser or equal priority. Inside an FCFS group requests of all classes are serviced in the first-in-first-out order. Assume also that all buffers for storing requests of different classes have infinite capacity and the queuing system utilization parameter

Y approaches 1 from below. Note that $Y = \sum_{h=1}^H y_h$, where $y_h = \lambda_h b_h$ is the load offered by requests of h -th class. In the described case (IBCQM) utilization parameter is the same as offered load factor because buffer capacity is infinite.

In [7], the described model is analytically solved in the form of Laplace transforms for the sojourn time. Respective transforms can be used to calculate any moment of sojourn time distribution. However, this solution can't be used if offered load is ≥ 1 . Using FBCQM is also hardly possible because there is no analytical solution for an NPP queuing with limited buffer capacity. Simulating such systems is extremely resource-consuming because of convergence issues. So in this paper we had to use a combined research approach to get analytical approximations for FBCQM and verify them using simulation for a limited amount of baseline cases.

3. Analysis of Queues with Infinite Buffer Capacity

Using the approach offered in [7], we calculate mean sojourn time u_h and its second moment $u_h^{(2)}$ for requests of h -th class as follows:

$$u_h = \frac{\sum_{i=1}^H \lambda_i b_i^{(2)}}{2(1 - R'_h)(1 - R''_h)} + b_h, \quad (1)$$

$$\begin{aligned}
u_h^{(2)} &= \frac{\sum_{i=1}^H \lambda_i b_i^{(3)}}{3(1 - R'_h)^2(1 - R''_h)} + \\
&+ \frac{\sum_{i=1}^H \lambda_i b_i^{(2)} \sum_{i=1}^H \lambda_i b_i^{(2)} (1 - 0, 5q_{hi}(3 - q_{hi}))}{2(1 - R'_h)^2(1 - R''_h)^2} + \\
&+ \frac{\sum_{i=1}^H \lambda_i b_i^{(2)} \sum_{i=1}^H \lambda_i b_i^{(2)} q_{ih}(3 - q_{ih})}{4(1 - R'_h)^3(1 - R''_h)} + \\
&+ \frac{b_h \sum_{i=1}^H \lambda_i b_i^{(2)}}{(1 - R'_h)(1 - R''_h)} + b_h^{(2)},
\end{aligned} \tag{2}$$

where $R'_h = \sum_{i=1}^H 0, 5q_{ih}(3 - q_{ih})\lambda_i b_i$ and $R''_h = \sum_{i=1}^H (1 - 0, 5q_{hi}(3 - q_{hi}))\lambda_i b_i$ are loads cumulatively offered by all requests having priority higher than h and by requests having priority not lesser than h respectively. Thus, $R = R_H^{(2)}$ represents the total utilization parameter or the total offered load.

Values u_h and $u_h^{(2)}$ can be used to calculate coefficient of variation (CV) of sojourn time of the requests of h -th class: $\nu_h = \sqrt{u_h^{(2)} - u_h^2}/u_h$. Average amount of requests inside the queuing system and second moment of that amount for class h can be calculated as follows (see [7]):

$$m_h = \lambda_h u_h; m_h^{(2)} = \lambda_h^2 u_h^{(2)} + \lambda_h u_h.$$

From that we can derive CV of requests amount of class h as $\nu_{m_h} = \sqrt{\nu_{u_h}^2 + 1/m_h}$. By using u_h and $u_h^{(2)}$ one can approximate distribution function of sojourn time to calculate jitter (delay variation).

4. Analysis of FCFS Queues

In order to narrow the range of varied parameters we first analyze non-priority case in which requests of all classes are serviced on a First-Come-First-Served basis. Suppose that distribution of service time is the same for all H classes in the sense of being defined by the same first three moments of the service time distribution: $b_h = b, b_h^{(2)} = b^{(2)}, b_h^{(3)} = b^{(3)}$.

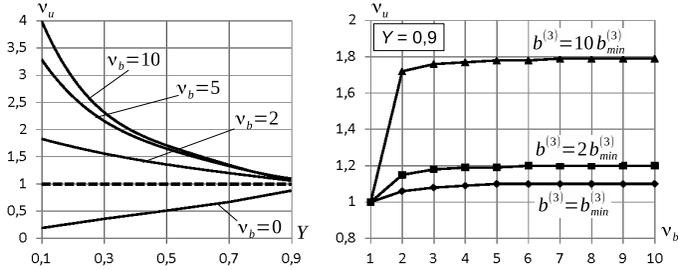


Figure 1. CV of sojourn time for FCFS queue

Sojourn time of all classes will also be the same given that all classes of requests arrive according to Poisson process.

Using (1) and (2) we analyzed how CV of service time $\nu_b = \sqrt{b^{(2)} - b^2}/b$ influences sojourn time average and CV ν_u . On Fig. 1 we show how ν_u is affected by total load Y for different values of service time CV. Presented plots suggest that ν_u depends only on utilization parameter, second and third moments of service time. It means ν_u doesn't depend on the first moment of service time (apart from its contribution to utilization). As utilization parameter approaches 1, the value of ν_u approaches 1 (dashed line marks values of ν_u for $\nu_b = 1$). For $Y = 0,98$ values of ν_u lie in the range from 0,88 to 1,1 for $\nu_b < 10$. In practice, any ν_u may be considered pretty close to 1 for $R > 0.9$.

Note that the third moment of service time may significantly influence ν_u . For example, for two-phase hyper-exponential distribution of service time, the third moment depends of the probability q of choosing one of the phases where $0 < q \leq \frac{2}{1+\nu_b^2}$ (see [8]). For $q = \frac{2}{1+\nu_b^2}$ the third moment reaches its minimum at $b^{(3)} = b_{min}^{(3)} = 1,5(1 + \nu_b^2)^2 b^3$. On Fig. 1 we also show how ν_u is influenced by $b^{(3)}$ for utilization parameter 0,9.

The plot also suggests that in a highly loaded system as ν_b increases ν_u is also increasing almost reaches its asymptotic maximum somewhere between $3 < \nu_b < 5$. A further increase of ν_b does not affect ν_u . For example, increasing $b^{(3)}$ ten-fold is unable to increase ν_u above 1,8 even for big values of ν_b . This enabled us limiting amount of further experiments by varying ν_b in the interval between 1 and 5.

5. Analysis of Priority Queuing Systems

Assume a queuing system with $H = 5$ classes in which smaller class index corresponds to a higher priority in the NPP queuing rule. Analyzing formulae (1) and (2) one can see that sojourn time average and its standard

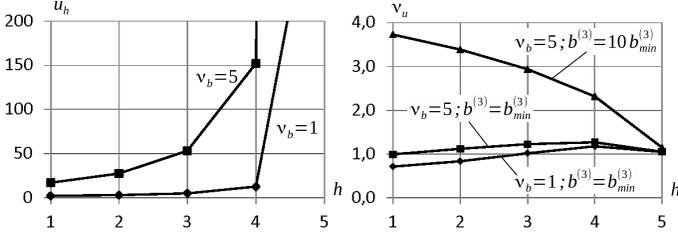


Figure 2. Average and CV of sojourn time for different classes

deviation (SD) for class h don't depend on inter-class priority assignment of all other classes. In other words, any q_{ij} ($i = 1, \dots, H; i \neq h; j = 1, \dots, H; j \neq h$) value doesn't affect h -th class performance metrics. An intuitive explanation for this fact is that the sum of service times of higher-priority requests standing in the queue doesn't depend on the order in which they get serviced.

Another point is that both sojourn time average and its SD increase as class index increases: we see that higher priority means lesser sojourn time average and SD (left plot on Fig. 2). However, CV of sojourn time is not always monotonous as may be seen on the right part of Fig. 2 where for $Y = 0,99$ the CV of service time significantly affects average sojourn time whereas it almost has no influence on CV of sojourn time. Note also that $b^{(3)}$ affects high priority classes more than it affects low priority classes.

Next we analyzed the influence of offered load structure on performance metrics. Suppose that total load is $Y = 0,95$ and per-class loads that constitute Y form a decreasing sequence: $y_1 = 0,8 > y_2 = 0,08 > y_3 = 0,04 > y_4 = 0,02 > y_5 = 0,01$. On Fig. 3 we see that as ν_b changes in the interval from 1 to 5, the values of ν_u are almost left unchanged: their relative difference is less than 3%.

On the contrary, changes of $b_h^{(3)}$ (with b_h and ν_b being fixed) may reverse the influence of priority on sojourn time CV: increasing priority may lessen ν_u for small $b_h^{(3)}$ and decrease it for bigger $b_h^{(3)}$. At the same time, the class with the smallest priority doesn't follow this pattern: $\nu_u|_{h=5}$ is almost not influenced by $b_5^{(3)}$. Besides that, when $b^{(3)} = b_{min}^{(3)}$, CV of sojourn time of the highest-priority class is much less than that of all other classes — this is because highest-priority class offers the biggest load.

Suppose now that per-class loads that constitute Y form an increasing sequence: $y_1 = 0,01 < y_2 = 0,02 < y_3 = 0,04 < y_4 = 0,08 < y_5 = 0,8$. The least-priority class offers the biggest load. As shown on Fig. 3, CV of sojourn time are almost not affected by CV of service time (just like in the previous case). Increasing third moment of service time ten-fold leads to a four-fold increase of sojourn time CV of the highest-priority class. At

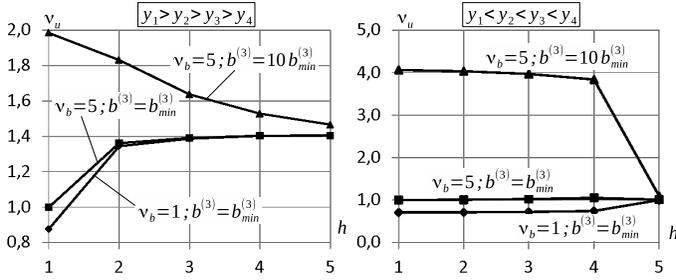


Figure 3. Sojourn time CV for different service time distributions

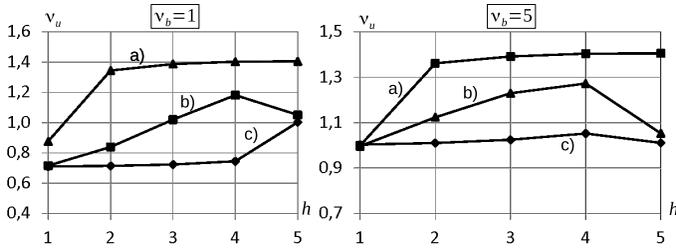


Figure 4. Sojourn time CV for varying load scenarios

the same time CV of the least-priority class approaches 1 independent of service time moments. Also note that classes 1-4 have almost identical CV of sojourn time — this is because respective classes offer a very modest load.

A generalization of load structure influence is shown on Fig. 4 for $\nu_b = 1$ and $\nu_b = 5$. Each plot reflects three scenarios: a) loads decrease as priority decreases: $y_1 > y_2 > \dots > y_5$; b) all loads are equal: $y_1 = y_2 = \dots = y_5$; c) loads increase as priority decreases: $y_1 < y_2 < \dots < y_5$. The total offered load is $Y = 0,99$ in all scenarios.

Scenarios a) and c) yield an upper and lower bound of sojourn time CV range respectively. Scenario b) yields a curve in between a) and c) as expected. As CV of service time is increased, the range of sojourn time CV decreases at the expense of the lower bound with upper bound almost left unchanged. An interesting point in scenario b) is that CV of sojourn time of the least-priority class is unexpectedly less than that of higher-priority classes. Besides that, least-priority class CV of sojourn time approaches 1.

6. Estimating Jitter in Priority Queuing Systems

Following the approach offered in [4], we define jitter (delay variation) of sojourn time J_{u_h} as upper bound for $(1 - \delta_h)$ -quantile of sojourn time probability distribution minus the minimal value of sojourn time (often considered 0). Upper bound of jitter $J_{u_h} \leq T_{u_h}$ can be determined using equation $F_{u_h}(T_{u_h}) = 1 - \delta_h$, where $F_{u_h}(\tau)$ is cumulative distribution function (CDF) of sojourn time, δ_h is the probability of sojourn time exceeding T_{u_h} . For example, for exponential distribution $F_{u_h}(T_{u_h}) = 1 - \exp(-T_{u_h}/u_h)$, so jitter is $T_{u_h} = -u_h \ln \delta_h$, where u_h is the average sojourn time. In computer networks $\delta_h = 10^{-3}$, so exponential jitter is $T_{u_h} \approx 6,91u_h$.

To estimate J_{u_h} the following two-moment approximation is possible using formulae (1) and (2). Values u_h and ν_u can be used to approximate CDF of $F_{u_h}(\tau)$ by first two moments using one of the multi-exponential distributions: 1) Erlang distribution for $\nu_u \leq 1$; or 2) hyper-exponential distribution for $\nu_u > 1$ (see [8]).

For $\nu_h \leq 1$, Erlang distribution with shape parameter k ($k = 1, 2, \dots$) approximates a chosen probability distribution by selecting k as follows: $k = \text{ceil}(1/\nu_h^2)$, where $\text{ceil}(x)$ returns the nearest integer not less than x . This allows to get an upper-bound approximation for T_{u_h} by solving the equation $F_{u_h}(T_{u_h}) = 1 - \delta_h$ for respective Erlang distribution. E.g. for $\delta_h = 10^{-3}$ and different k we get $T_{u_h} \approx 4,6u_h$ (for $k = 2, \nu_u \approx 0,71$), $T_{u_h} \approx 3,74u_h$ (for $k = 3, \nu_h \approx 0,58$), $T_{u_h} \approx 2,27u_h$ (for $k = 4, \nu_h = 0,5$) and so on.

For $\nu_h > 1$, hyper-exponential distribution approximates a chosen probability distribution using the following CDF: $F_h(\tau) = 1 - qe^{-\tau/t}$, where $q = \frac{2}{1+\nu_h^2}$ and $t = \frac{1+\nu_h^2}{2}u_h$ (this turns out to be the Cox distribution which is a special case of hyper-exponential distribution). After solving the equation $F_{u_h}(T_{u_h}) = 1 - \delta_h$, we get this (see also [8]):

$$T_{u_h} = -0,5(1 + \nu^2) \ln(0,5\delta_h(1 + \nu^2))u_h. \quad (3)$$

Offered analytical model allows to estimate average sojourn time u_h and its jitter $J_{u_h} \leq T_{u_h}$, but doesn't allow to estimate probability loss π_h for a given class h in an FBCQM. Exact analytical solution is only possible for queuing systems that can be represented by a Markov process. In general case, however, only rough estimation methods have to be applied. We offer one such method that is based on estimating buffer capacity of FBCQM using an existing analytical model for IBCQM in which we calculate such queue size that is not exceeded with a given probability equal to π_h . Resulting queue size is then used as an approximation for buffer capacity in FBCQM that ensures required loss probability π_h .

7. Estimating Buffer Capacity in Priority Queuing Systems

The offered method implies two-moment approximation of amount of requests CDF. As before, Erlang and hyper-exponential distributions are used. Doing so we have to interpret the amount of requests as a continuous random variable (whereas it is inherently discrete) because existing discrete distribution functions (e.g. Negative Binomial) don't allow to solve the problem explicitly.

In case of the Cox distribution the following formulae (analogous to (3)) was derived to estimate the maximal amount of requests that can be exceeded with probability less than π_h :

$$M_h = ceil(-0,5(1 + \nu_{m_h}^2) \ln(0,5\pi_h(1 + \nu_{m_h}^2))m_h).$$

So if buffer capacity E_h for class h is greater than or equal to M_h , then we expect loss probability for class h to not exceed π_h . However, this is a rough two-moment approximation. In order to use three-moment approximation one may have to use normal-distribution approximation proposed in paragraph 8.2.4 of [4]. To do so, however, one would have to derive a formula for the third moment of sojourn time from [7] by analogy with formulae (1) and (2).

8. A numerical example

Effectiveness and correctness of the offered approach of estimating priority-queuing metrics was confirmed by multiple simulation experiments using GPSS World environment. We present one such experiment here.

Consider a queuing system with 5 classes of requests each arriving according to Poisson process with rates 0,05; 0,025; 0,01; 0,008 and 0,005 per second. Average service time is the same for all classes: 10 s. Total offered load is thus 0,98. CV of service time is 0,5; 0; 0,7; 1 and 2 for respective classes. Third moments of service time are 1000; 1000; 1000; 6000 and 37500 s³ respectively. NPP queuing discipline is used with the following priority-assignment rule: higher class index corresponds to lower priority.

Table 1 is used to present performance metrics that were obtained using three different methods under assumption that the following loss probability constraint is used: $\pi_h \leq 0,001 = 0,1\%$. Method 1 is based on the offered approach described in previous sections. Method 2 is based on using Chebyshev's inequality to estimate upper bound of respective performance metrics. Method 3 is based on simulation in GPSS World environment and is aimed at verifying correctness of Method 1 (500 million transactions were used to get the result so confidence interval is too narrow to be shown).

In table 1 one can see that loss probabilities calculated by Method 1 for each class are guaranteed to be less than the chosen constraint for π_h . Note

that Method 1 allows to improve the accuracy of buffer capacity estimation up to several times as compared to Method 2. This conclusion is not weakened if Vysochanskij–Petunin inequality (extension of Chebyshev’s inequality for unimodal distributions) is used to estimate buffer capacity in Method 2.

9. Conclusions

The purpose and main novelty of the paper was to offer an analytical solution for QoS-based computer network metrics of a priority queuing model in heavy traffic conditions. Problems of simulating highly-loaded priority queuing systems require development of analytical methods to calculate different performance metrics. As a result, several analytical approximations were offered to estimate jitter (delay variation), average sojourn time and required buffer capacity for heavy traffic case. Besides that, an exact solution was developed to calculate first and second moment of sojourn time.

Table 1
Comparison of Simulation Results with Proposed Approximations

Performance metric	Requests class index				
	1	2	3	4	5
Average sojourn time (s)	24,3	67,4	201,2	692,9	5131,4
Average amount of requests	1,22	1,68	2,01	5,54	25,66
CV of sojourn time	0,60	0,92	1,15	1,19	1,23
CV of amount of requests	1,09	1,20	1,35	1,27	1,24
Sojourn time jitter (s)	120,6	435,8	1583,7	5634,4	42986,2
Buffer capacity E_h for $\pi_h = 0,1\%$ according to proposed approximations	10	14	19	48	218
Buffer capacity for $\pi_h = 0,1\%$ according to Chebyshev’s inequality	42	64	86	222	1009
Loss probability (%) according to simulation for buffer capacity E_h	0,041	0,071	0,082	0,058	0,026

References

1. *Vishnevsky V., Semenova O.* Polling Systems: Theory and Applications for Broadband Wireless Networks. — Academic Publishing, London. — 2012. — 317 p.
2. *Bogatyrev V.A.* Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer // Automatic Control and Computer Sciences. — 1999. — No. 33 (1). — P. 57–63.
3. *Bogatyrev V.A., Golubev I.Yu.* Service model of inhomogeneous flow at priority processing duplicate criticality queries // Journal of computer & information technology. — 2014. — No. 4. — P. 27–32.
4. Network performance objectives for IP-based services. // International Telecommunication Union Recommendation ITU-T Y.1541. — Geneva, 2012. — 66 p.
5. *Gautam N.* Analysis of Queues: Methods and Application. — CRC Press, 2014. — 762 p.
6. *Aliev T. I.* Osnovy modelirovaniya diskretnykh sistem. — ITMO University, Saint-Petersburg. — 2009. — 363 p. — ISBN 978-5-7577-0336-7.
7. *Aliev T. I., Maharevs E.* Queuing Disciplines Based on Priority Matrix // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. — 2014. — Vol. 14. — No. 6 (94). — P. 91–97.
8. *Aliev T. I.* Approximation of Probability Distributions in Queuing Models // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. — 2013. — Vol. 13. — No. 2 (84). — P. 88–93.

УДК 519.217

Моделирование технологии NB-IoT для приложений интернета вещей

В. О. Бегишев*, В. И. Петров^{†‡}, А. К. Самуйлов^{†*},
Д. А. Молчанов^{†*}, Ю. В. Гайдамака*

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

† *Департамент электроники и телекоммуникаций,
Технологический университет,
ул. Корккянкюллюнкату, д.10, г.Тампере, Финляндия, 33720*

‡ *Кафедра сетей связи и передачи данных,
Санкт-Петербургского государственного университета
телекоммуникаций им. проф. М.А.Бонч-Бруевича,
пр. Большевиков, д. 22Б корп. 1, Санкт-Петербург, Россия, 193232*

Аннотация. С внедрением стандарта NB-IoT операторы традиционной сотовой связи смогут подключить к своим сетям сотни тысяч IoT-устройств. Предполагается, что сети NB-IoT найдут широкое применение в решениях для умных городов, умной утилизации мусора и многих других задач. В работе предложена аналитическая модель для анализа основных показателей эффективности функционирования беспроводной сети LTE с поддержкой NB-IoT, а именно, вероятности потери сообщения и среднего значения времени задержки сообщения в системе. Для модели представлены результаты численного эксперимента для сот с радиусом покрытия 10 и 30 километров. Предложенная математическая модель может быть использована для анализа характеристик рассматриваемой технологии Интернета вещей в телекоммуникационных сетях.

Ключевые слова: Интернет вещей, NB-IoT (Narrowband IoT), модель системы, LTE (Long Term Evolution).

1. Введение

Низкое энергопотребление, доступная цена датчика, высокая дальность передачи радиосигнала, а также поддержка большого числа соединений являются основными требованиями к беспроводной сети с поддержкой Интернета вещей (Internet of Things). Для таких требований разработан новый тип сетей NB-IoT, для которой процесс стандартизации завершен (в июне 2016 года) и эта технология официально включена в LTE Rel. 13 [1, 2]. NB-IoT может быть относительно легко внедрена в существующие сети сотовой связи LTE, добавляя несколько новых категорий устройств, разработанных специально с целью уменьшения сложности оконечных устройств. Чтобы достичь необходимой эффективности, ширина беспроводного канала была снижена до 180 кГц, что позволило получить усиление сигнала на 20 дБ. Данные изменения позволяют значительно увеличить эффективную дальность

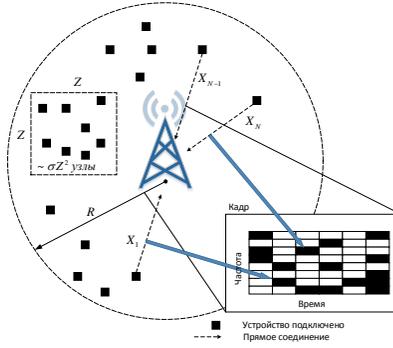


Рис. 1. Модель системы NB-IoT

передачи небольших объемов данных для маломощных устройств и повысить их автономность работы от батареи до 10 и более лет в зависимости от сценария применения.

В работе проведен анализ производительности технологии NB-IoT с точки зрения вероятностных характеристик обслуживания, таких как вероятность потери сообщения, распределение и средняя задержка сообщения в системе. Представленная модель позволяет определить диапазоны нагрузок при которых выполняются заданные параметры качества обслуживания устройств.

2. Модель системы

Рассмотрим сценарий, представленный на Рис. 1. Предположим, что зона обслуживания базовой станции имеет форму окружности с радиусом R с равномерно распределёнными в ней N устройствами. Плотность датчиков на 1 км^2 определяется формулой $\sigma = N/\pi R^2$. Интенсивность поступления заявки от одного устройства будем обозначать λ_t . Тогда $\lambda_I = \lambda_t \Delta$ интенсивность поступления сообщений от одного устройства за слот, где Δ – длительность одного кадра LTE. Датчики отправляют сообщение размером 100 бит. Таким образом, для передачи любого сообщения требуется один ресурсный блок NB-IoT.

Анализ описанной в разделе системы проведен с помощью системы массового обслуживания в дискретном времени. Суммарный поток поступления сообщений от всех датчиков зоне обслуживания соты является пуассоновским с интенсивностью $\lambda = N\lambda_t$. Под NB-IoT выделяется $K = 6 \times 12 = 72$ виртуальных канала в одном слоте. В итоге моделируется система массового обслуживания в дискретном времени с

групповым поступлением и обслуживанием сообщений. Длительность обслуживания каждой заявки постоянна.

Сообщения поступают в систему в момент времени непосредственно перед окончанием слота. Обслуживаются заявки группами размера не более, чем штук, окончание обслуживания группы происходит в момент окончания слота. Состояние системы рассматривается в момент времени непосредственно после окончания обслуживания очередной группы. При этом вновь пришедшие заявки не могут моментально поступить на обслуживание в слоте, в котором они пришли, а будут обслужены только в следующем. Группа из i сообщений, поступившая в систему, в которой уже имеется k сообщений, принимается полностью, если $k + i \leq L$, в противном случае $k + i - L$ сообщений из группы теряется.

Для описанной системы анализируются распределение вероятностей потери i сообщений, среднее значение времени задержки сообщения в системе. Данные значения позволяют выбирать плотность сенсоров таким образом, чтобы обеспечивались требования по доставке сообщения в пределах заданного интервала времени. Также следует отметить, что задержка неявно влияет на энергоэффективность работы сенсора, определяя среднее время нахождения в режиме повышенного энергопотребления пока сенсор ожидает начала передачи данных.

3. Заключение

В представленной работе, используя аппарат теории очередей в дискретном времени, построена модель обслуживания сообщений технологии NB-IoT. Для предложенной модели получены вероятностные характеристики обслуживания, включая распределение количества потерянных сообщений, вероятность потери сообщения, распределение и среднее время обслуживания одного сообщения. Приведены численные результаты для сот радиуса 10 и 30 километров. Предложенная модель может быть использована для определения максимального допустимого количества сенсоров в зоне покрытия одной соты при заданных параметрах качества обслуживания.

Благодарности

Исследование выполнено в рамках соглашения № 02.а03.21.0008 от 24.04.2016 г. между Министерством образования и науки Российской Федерации и Российским университетом дружбы народов. Работа поддержана грантами РФФИ № 15-07-03051, 15-07-03608, 16-07-00766.

Литература

1. *3GPP*, “Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)”, 3GPP TR 45.820/r13, December 2015.
2. *3GPP*, “Standardization of NB-IoT completed”, http://www.3gpp.org/news-events/3gpp-news/1785-nb_iot_complete, June 2016.
3. *G. Cardone, A. Corradi, L. Foschini and R. Ianniello*, ParticipAct: A Large-Scale Crowdsensing Platform. “IEEE Transactions on Emerging Topics in Computing” — Vol.4, — P. 21–32 —September 2015.
4. *B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou*, “Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm”, ACM Computing Surveys — Vol.48, — P. 7:1–7:31, —January 2016.
5. *Moustafa, Hassnaa and Zhang*, Vehicular Networks: Techniques, Standards, and Applications, Boston, MA, USA: Auerbach Publications, 1st ed., 2009.

UDC 519.217

Modeling NB-IoT technologies for the Internet of Things Applications

V. O. Begishev*, **V. I. Petrov^{†‡}**, **A. K. Samuylov^{†*}**,
D. A. Moltchanov^{†*}, and **Yu. V. Gaidamaka***

[†] *Department of Electronics and Communications Engineering
Tampere University of Technology*

Korkeakoulunkatu str. 10, Tampere, 32720, Finland

^{*} *Department of Applied Probability and Informatics
RUDN University*

Miklukho-Maklaya str. 6, Moscow, 117198, Russia

[‡] *Department of Telecommunication Networks and Data Transmission*

The Bonch-Bruевич Saint-Petersburg State University of Telecommunications

Prospekt Bolshhevikov 22 B, St. Petersburg, 193232, Russia

In this paper we develop an analytical model for recently standardized NB-IoT technology. For the proposed model we derive the probability of message loss as well as the distribution and the mean value of the message delay in the system. We then proceed illustrating the numerical results for cells of different sizes. The proposed model can be used as a baseline for analysis of the NB-IoT technology deployments and its various improvements.

Keywords: Internet of Things, NB-IoT (Narrow Band IoT), system model, LTE (Long Term Evolution).

УДК 517.9

Стационарные решения уравнений Власова для высокотемпературной двухкомпонентной плазмы

Ю. О. Беляева*, А. Л. Скубаческий*

** Кафедра прикладной математики,
Российский университет дружбы народов,
ул. Миклухо-Макляя, д.6, Москва, Россия, 117198*

Аннотация. Рассматривается первая смешанная задача для системы уравнений Власова–Пуассона в бесконечном цилиндре, описывающая эволюцию плотностей распределения ионов и электронов в высокотемпературной плазме при наличии внешнего магнитного поля. Построены стационарные решения с тривиальным потенциалом самосогласованного электрического поля с носителями плотностей распределения заряженных частиц во внутреннем цилиндре.

Ключевые слова: уравнений Власова–Пуассона, смешанная задача, стационарные решения, однородное магнитное поле.

1. Введение

Уравнения Власова, или кинетические уравнения с самосогласованным полем, впервые были получены в 1938г в работе [3]. В настоящее время они представляют одну из наиболее известных математических моделей кинетической теории газов. Популярность этих уравнений связана прежде всего с многочисленными приложениями, среди которых наибольший интерес вызывает использование уравнений Власова в изучении высокотемпературной, разреженной плазмы и прежде всего процессов управления термоядерным синтезом. Глобальная разрешимость “сглаженных” уравнений Власова исследована в работах Р.Л. Добрушина и В.П. Маслова [6, 8]. Существование и устойчивость слабых решений начальных и начально-краевых задач для системы уравнений Власова–Пуассона и Власова–Максвелла изучались в работах А.А. Арсеньева, В. В. Козлова, Р.Дж. ДиПерна и П.Л. Лионсом, Ю. Веклером [1, 7, 5, 12] и др. Вопрос о существовании глобальных классических решений задачи Коши для уравнений Власова–Пуассона исследовался в работах Ю. Батта, К. Пфаффельмозера и Дж. Шеффера [2, 9, 10] и ряде других работ. Существование классических решений начально-краевых задач для уравнений Власова–Пуассона исследовано в значительно меньшей степени. В работе Я. Гуо [4] доказано существование глобального классического решения второй начально-краевой задачи для системы уравнений Власова–Пуассона в полупространстве. А.Л. Скубачевским [11] был получен результат о существовании и единственности классического решения в некоторой окрестности стационарного решения с носителями плотностей распределения заряженных частиц, лежащими на некотором

расстоянии от цилиндрической границы. В представленной работе построены примеры стационарных решений с тривиальным потенциалом самосогласованного электрического поля.

2. Основная часть

Рассмотрим систему уравнений Власова–Пуассона в бесконечном цилиндре

$$-\Delta\varphi(x, t) = 4\pi e \int_{\mathbb{R}^3} \sum_{\beta} \beta f^{\beta}(x, v, t) dv \quad (x \in Q, 0 < t < T), \quad (1)$$

$$\frac{\partial f^{\beta}}{\partial t} + (v, \nabla_x f^{\beta}) + \frac{\beta e}{m_{\beta}} \left(-\nabla_x \varphi + \frac{1}{c} [v, B], \nabla_v f^{\beta} \right) = 0 \quad (2)$$

$$(x \in Q, v \in \mathbb{R}^3, 0 < t < T, \beta = \pm 1)$$

с начальными условиями

$$f^{\beta}(x, v, t)|_{t=0} = f_0^{\beta}(x, v) \quad (x \in \bar{Q}, v \in \mathbb{R}^3, \beta = \pm 1) \quad (3)$$

и краевым условием Дирихле

$$\varphi(x, t) = 0 \quad (x \in \partial Q, 0 \leq t < T). \quad (4)$$

Здесь $Q = G \times \mathbb{R}$, $G \subset \mathbb{R}^2$ — ограниченная область с границей $\partial G \in C^{\infty}$, $\partial Q = \partial G \times \mathbb{R}$, $f^{\beta} = f^{\beta}(x, v, t)$ — функция плотности распределения положительно заряженных ионов, если $\beta = +1$, и электронов, если $\beta = -1$, в точке x со скоростью v в момент времени t ; $\varphi = \varphi(x, t)$ — потенциал самосогласованного электрического поля; ∇_x и ∇_v — градиенты по x и v , соответственно; m_{+1} и m_{-1} — массы иона и электрона; e — заряд электрона; c — скорость света; B — индукция внешнего магнитного поля; (\cdot, \cdot) — скалярное произведение в \mathbb{R}^3 ; $[\cdot, \cdot]$ — векторное произведение в \mathbb{R}^3 .

Пусть $B_{\rho}(x_0) = \{x \in \mathbb{R}^3: |x - x_0| < \rho\}$, и пусть $B_{\rho} = B_{\rho}(0)$, тогда $|B_{\rho}| = 4\pi\rho^3/3$.

Для определения классического решения задачи (1)–(4), введем некоторые функциональные пространства. Обозначим через $C^s(\mathbb{R}^n)$ ($C^s(\Omega)$), $s \geq 0$, $n \in \mathbb{N}$, пространство Гёльдера непрерывных функций в \mathbb{R}^n (Ω), имеющих все непрерывные производные в \mathbb{R}^n (Ω) вплоть до k -го порядка, $k = [s]$, с конечной нормой

$$\|u\|_s = \max_{|\alpha| \leq k} \sup_x |\mathcal{D}^{\alpha} u(x)|, \quad \text{если } s = k \in \mathbb{Z}, 0 \leq k,$$

$$\|u\|_s = \|u\|_k + |u|_{\sigma}, \quad \text{если } s = k + \sigma, 0 \leq k \in \mathbb{Z}, 0 < \sigma < 1,$$

где $\Omega \subset \mathbb{R}^n$ — область с границей $\partial\Omega \in C^\infty$ или цилиндр $Q = G \times \mathbb{R} \subset \mathbb{R}^3$,

$$|u|_\sigma = \max_{|\alpha|=k} \sup_{x \neq y} |x - y|^{-\sigma} |D^\alpha u(x) - D^\alpha u(y)|,$$

$$D^\alpha = \left(\frac{\partial}{\partial x_1} \right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n} \right)^{\alpha_n}, \quad \alpha = (\alpha_1, \dots, \alpha_n), \quad |\alpha| = \alpha_1 + \dots + \alpha_n.$$

Положим $C(\mathbb{R}^n) = C^0(\mathbb{R}^n)$ и $C(\bar{\Omega}) = C^0(\bar{\Omega})$. Аналогично мы можем ввести пространство $C^1(\bar{Q} \times \mathbb{R}^3 \times [0, T])$, состоящее из непрерывных, ограниченных функций, имеющих непрерывные, ограниченные производные первого порядка в $\bar{Q} \times \mathbb{R}^3 \times [0, T]$.

Определение 1. Вектор-функцию $\{\varphi, f^\beta\}$, $\varphi \in C([0, T], C_0^{2+\sigma}(\bar{Q}))$, $f^\beta \in C^1(\bar{Q} \times \mathbb{R}^3 \times [0, T])$, мы назовем классическим решением задачи (1)–(4), если $\{\varphi, f^\beta\}$ удовлетворяет уравнениям (1), (2), начальным условиям (3) и краевому условию (4).

Важную роль в исследовании уравнений Власова играют стационарные решения.

Определение 2. Вектор-функцию $\{\dot{\varphi}, \dot{f}^\beta\}$, $\dot{\varphi} \in C_0^{2+\sigma}(\bar{Q})$, $\dot{f}^\beta \in C^1(\bar{Q} \times \mathbb{R}^3)$, мы назовем стационарным решением уравнений (1), (2) с краевым условием (4), если $\{\dot{\varphi}, \dot{f}^\beta\}$ удовлетворяет уравнениям

$$-\Delta \dot{\varphi}(x) = 4\pi e \int_{\mathbb{R}^3} \sum_{\beta} \beta \dot{f}^\beta(x, v) dv \quad (x \in Q),$$

$$(v, \nabla_x \dot{f}^\beta) + \frac{\beta e}{m_\beta} \left(-\nabla_x \dot{\varphi} + \frac{1}{c} [v, B], \nabla_v \dot{f}^\beta \right) = 0 \quad (x \in Q, v \in \mathbb{R}^3, \beta = \pm 1)$$

и краевому условию (4).

Сформулируем теперь условия, которым должны удовлетворять магнитное поле B и начальные распределения плотностей заряженных частиц $f_0^\beta(x, v)$. Обозначим $G_\delta = \{x' \in G : \text{dist}(x', \partial G) > \delta\}$, $Q_\delta = \{x \in Q : \text{dist}(x, \partial Q) > \delta\}$, где $\delta > 0$. Предполагая, что $G_{2\delta} \neq \emptyset$, обозначим через $\delta_0 = \delta_0(\delta) > 0$ наибольший радиус круга, вписанного в $G_{2\delta}$.

Условие 1. Пусть $B = (0, 0, h)$ для $x \in \bar{Q}$, где $h > 0$ не зависит от x и

$$32 \frac{c\rho m_{+1}}{e\delta} < h.$$

Условие 2. Пусть $f_0^\beta \in C^\infty(\bar{Q} \times \mathbb{R}^3)$ — неотрицательные функции, и пусть $\text{supp} f_0^\beta \subset Q_{2\delta} \times B_{\rho/4}$, где $\rho > 0$, $\delta_0 > \delta$.

В работе [12] был получен следующий результат:

Теорема 1. Пусть $\delta > 0$ таково, что $G_{2\delta} \neq \emptyset$ и $\delta_0 > \delta$, и пусть для этого δ и некоторых $h, \rho > 0$ выполняется условие 1. Тогда для любого $\alpha > 0$ существует стационарное решение уравнений с краевым условием (4) $\{0, \mathring{f}^\beta\}$, обладающее следующими свойствами: $\mathring{f}^\beta \in C^\infty(\bar{Q} \times \mathbb{R}^3)$, $\text{supp } \mathring{f}^\beta \subset Q_{2\delta} \times B_{\rho/4}$.

Основываясь на теореме 1 и используя метод неопределенных коэффициентов получен результат:

Следствие 1. Пусть выполнены условия теоремы 1, тогда функции $\{0, \mathring{f}_1^\beta, \mathring{f}_2^\beta\}$ и $\{0, \mathring{f}_1^\beta, \mathring{f}_3^\beta\}$, где

$$\mathring{f}_1^\beta(x, v) = \psi_1^\beta(|v|^2),$$

$$\mathring{f}_2^\beta(x, v) = \psi_3^\beta \left(\left(\frac{\beta e h}{c m_\beta} x_1 + \beta v_2 \right)^4 + \left(\frac{\beta e h}{c m_\beta} x_2 - \beta v_1 \right)^4 \right),$$

$$\mathring{f}_3^\beta(x, v) = \psi_3^\beta \left(\left(\frac{\beta e h}{c m_\beta} x_1 + \beta v_2 \right)^{2m} + \left(\frac{\beta e h}{c m_\beta} x_2 - \beta v_1 \right)^{2m} \right),$$

являются стационарными решениями системы уравнений (1)-(2), удовлетворяющими условию (4) и обладают свойствами: $\mathring{f}^\beta \in C^\infty(\bar{Q} \times \mathbb{R}^3)$, $\text{supp } \mathring{f}^\beta \subset Q_{2\delta} \times B_{\rho/4}$. Здесь $\mathring{f}_1^\beta(x, v)$, $\mathring{f}_2^\beta(x, v)$, $\mathring{f}_3^\beta(x, v)$ - неотрицательные четные функции, такие, что $\psi_1^\beta, \psi_2^\beta, \psi_3^\beta \in C^\infty(\mathbb{R})$, $\psi_1^\beta(0) = \psi_2^\beta(0) = \psi_3^\beta(0) = 1$, $\text{supp } \psi_1^{-1} \subset (-\rho_1^2, \rho_1^2)$, где $\rho_1 = (15\rho\delta_0/\delta)$, $\text{supp } \psi_2^{-1} \subset (-\rho_2^2, \rho_2^2)$, где $\rho_2 = (15\rho\delta_0/\delta)^2$, $\text{supp } \psi_3^{-1} \subset (-\rho_3^2, \rho_3^2)$, где $\rho_3 = (15\rho\delta_0/\delta)^m$.

Литература

1. Арсеньев А.А. ЖВМиМФ. 1975. Т. 15. № 1. С. 136–147.
2. Batt J. J. Different. Equat. 1977. V. 25. P. 342–364.
3. Власов А.А. О вибрационных свойствах электронного газа, ЖЭТФ, 8:3 (1938), 291–318.
4. Guo Y. Indiana Univ. Math. J. 1994. V. 43. P. 255– 320.
5. Di Perna R., Lions P.L. Communs Pure and Appl. Math. 1989. V. 42. P. 729–757.
6. Добрушин Р.Л. Функцион. анализ и его прил. 1979. Т. 13. В. 2. С. 48–58.
7. Козлов В.В. УМН. 2008. Т. 63. В. 4(382). С. 93–130.
8. Маслов В.П. В кн.: Современные проблемы математики. М.: ВИНТИ, 1978. Т. 11. С. 153–234.
9. Pfaffelmoser K.J. Different. Equat. 1992. V. 95. P. 281–303.

10. *Schaffer J.* Commun. Part. Different. Equats. 1991. V. 16. P. 1313–1335.
11. *Skubachevskii A.L.* Russian Mathematical Surveys. 2014. 69 (2), pp. 291–330.
12. *Weckler J.* Arch. Ration. Mech. and Anal. 1995. V. 130. P. 145–161.

UDC 517.9

Stationary solutions of the Vlasov equations for a two-component high temperature plasma.

J. O. Belyaeva*, A. L. Skubachevskii*

** Department of Applied Mathematics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

We consider the first mixed problem for the Vlasov-Poisson equations in an infinite cylinder. This problem describes the evolution of the density distribution of ions and electrons in a high temperature rarefied plasma under an external magnetic field. In the present paper we construct stationary solutions of the Vlasov-Poisson equations for a two-component plasma with trivial potential of the self-consistent electric field.

Keywords: Vlasov-Poisson equations, mixed problem, stationary solutions, homogeneous magnetic field.

UDC 519.711.2

Transport systems analysis using neural networks

A. I. Blinov, L. A. Sevastianov, S. A. Vasilyev

** Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

Abstract. Transport systems was studied using neural networks technology. Neural networks (NNs) capability of mapping the nonlinear functions of variables describing the behaviour of objects and the simplicity of designing their configuration favours their applications in transport. This paper presents representative examples in the scope of prediction of road traffic parameters, road traffic control, measurement of road traffic parameters, driver behaviour and autonomous vehicles, and transport policy and economics. The features of the solutions are examined. The review shows that feedforward multilayer neural networks are the most often utilised configurations in transportation research. No systematic approach is reported on the optimisation of the NN configurations to achieve a set level of performance in solving modelling tasks.

Keywords: transportation problem, economical process simulation; dynamics of complicated systems, neural networks.

1. Introduction

The solution of problems of mathematical simulation of complex transport systems at this stage is more difficult because of the large amount of data that must be analyzed [9]. For example, a huge number of possible options of traffic makes it difficult to obtain sufficient economical plan through empirical or expert. Application of mathematical methods and computing in transport planning gives considerable economic benefit. The transportation problem can be solved by simplex method but the matrix of the constraints of the transportation problem is often so complex that its solution developed special methods. These methods, like the simplex method, allows to find the initial reference solution and then improving it to an optimal solution, but this approach is costly.

The recent research of the dynamics of complex systems using neural networks. A model for a large network of "neurons" with a graded response (or sigmoid input-output relation) was studied [5]. The idea was used in biological systems was given added credence by the continued presence of such properties for more nearly biological "neurons". In the paper [2] [3] was given existence and uniqueness results for the equations describing the dynamics of some neural networks for which there were infinitely many cells. Such system was considered and neural nets which were modelled were described by the infinite system of ordinary differential equations.

Haiying Huang, Qiaosheng Du and Xibing Kang [6] studied a class of neutral high-order stochastic Hopfield neural networks with Markovian

jump parameters and mixed time delays. In paper [11] Xiao Liang, Linshan Wang, Yangfan Wang and Ruili Wang focused on the long time behavior of the mild solution to delayed reaction-diffusion Hopfield neural networks (DRDHNNs) driven by infinite dimensional Wiener processes. They analyzed the existence, uniqueness, and stability of this system under the local Lipschitz function by constructing an appropriate Lyapunov-Krasovskii function and utilizing the semigroup theory.

In the paper Hanihossein Abolfazli, Seyed Masoud Asadzadeh, Salman Nazari-Shirkouhi, Seyed Mohammad Asadzadeh, Kamran Rezaie [1] presented the application of time series and artificial neural network for improvement of energy forecasting in rail transport section. Sustekova D., Knutelska M. [10] investigated possibility of using neural networks in the road transport, the advantages and disadvantages of using neural networks, brief description of neural networks — definition and basic elements. Rahman S.A., Ansari M.S., Moinuddin A.A. [8] explored a problem of minimizing a linear cost function subject to linear constraints. They employed non-linear feedback, in the form of unipolar comparators, to introduce transcendental terms in the energy function ensuring fast convergence to the solution. The proof of validity of the energy function was also provided. The hardware complexity of the proposed circuit compared favorably with other proposed circuits for the same task.

In this paper [7] a modified Hopfield network was developed for solving linear programming problems. The internal parameters of the network was obtained using the valid-subspace technique. A complexity analysis between the main neural networks used in linear programming was also developed.

In this work, based on detailed analysis of the basic methods of numerical simulation of the transport task using the technology of neural networks for the transport systems built a mathematical model taking into account the cost. This paper presents a recurrent neural circuit for solving linear programming problems using methods from [7].

2. The Modified Hopfield Network

Besides providing a new approach for solving linear programming problems, artificial neural networks provide a method for exploring intrinsically parallel and adaptive processing architectures. In this paper, a modified Hopfield network with equilibrium points representing the problem solution has been developed. As introduced in [5], Hopfield networks are single-layer networks with feedback connections between nodes. In the standard case, the nodes are fully connected, i.e., every node is connected to all others nodes, including itself. The node equation for the continuous-time

network with n -neurons is given by:

$$\begin{cases} \dot{u}_i(t) = -\eta u_i(t) + \sum_{j=1}^n T_{ij} \nu_j(t) + I_i^b, \\ \nu_j(t) = g(u_i(t)), \end{cases}$$

where $u_i(t)$ is the current state of the i -th neuron; $\nu_j(t)$ is the output of the j -th neuron; I_i^b is the offset bias of the i -th neuron; $\eta u_i(t)$ is a passive decay term; and T_{ij} is the weight connecting the j -th neuron to i -th neuron; $g(u_i(t))$ is a monotonically increasing threshold function that limits the output of each neuron to ensure that network output always lies in or within a hypercube. It is shown in [5] that the equilibrium points of the network correspond to values $\nu_j(t)$ for which the energy function $E(t)$ associated with the network is minimized:

$$E(t) = -\frac{1}{2} \nu^T T \nu - \nu^T I^b.$$

The mapping of linear programming problems using a Hopfield network consists of determining the weight matrix T and the bias vector I^b to compute equilibrium points. A modified energy function $E_{mod}(t)$ is used here, defined as follows:

$$E_{mod}(t) = E_{conf}(t) + E_{opt}(t),$$

where $E_{conf}(t)$ is a confinement term that groups all the constraints imposed by the problem, and $E_{opt}(t)$ is an optimization term that conducts the network output to the equilibrium points. Thus, the minimization of $E_{mod}(t)$ of the modified Hopfield network is conducted in two stages:

i) minimization of the term $E_{conf}(t)$:

$$E_{conf}(t) = -\frac{1}{2} \nu^T T_{conf} \nu - \nu^T I_{conf},$$

where $\nu(t)$ is the network output, T_{conf} is weight matrix and I_{conf} is bias vector belonging to E_{conf} . This corresponds to confinement of $\nu(t)$ into a valid subspace that confines the inequality constraints imposed by the problem.

ii) minimization of the term $E_{opt}(t)$:

$$E_{opt}(t) = -\frac{1}{2} \nu^T T_{opt} \nu - \nu^T I_{opt}, \quad (1)$$

where T_{opt} is weight matrix and I_{opt} is bias vector belonging to $E_{opt}(t)$. This moves $\nu(t)$ towards an optimal solution (the equilibrium points). Thus, the operation of the modified Hopfield network consists of three main steps.

Step 1. Minimization of $E_{conf}(t)$, corresponding to the projection of $\nu(t)$ in the valid subspace:

$$\nu(t) = T_{conf}\nu(t) + I_{conf}, \quad (2)$$

where T_{conf} is a projection matrix ($T_{conf}T_{conf} = T_{conf}$) and $T_{conf}I_{conf} = 0$. This operation corresponds to an indirect minimization of E_{conf} .

Step 2. Application of a nonlinear symmetric ramp activation function constraining $\nu(t)$ in a hypercube:

$$d_i(\nu_i) = \begin{cases} L_{inf}, & \text{if } L_{inf} > \eta_i, \\ \eta_i, & \text{if } L_{inf} \leq \eta_i \leq L_{sup}, \\ L_{sup}, & \text{if } L_{sup} < \eta_i, \end{cases}$$

where $\nu(t) \in [L_{inf}, L_{sup}]$.

Step 3. Minimization of $E_{opt}(t)$, which involves updating of $\nu(t)$ in direction to an optimal solution (defined by T_{opt} and i_{opt}) corresponding to network equilibrium points, which are the solutions for the linear programming problems, by applying the gradient in relation to the energy term E_{opt} .

$$\frac{d\nu(t)}{dt} = \dot{\nu} = -\frac{\partial E_{opt}(t)}{\partial \nu},$$

$$\Delta\nu(t) = -\Delta E_{opt}(\nu)\Delta t = (T_{opt}\nu(t) - I_{opt})\Delta t.$$

Therefore, minimization of E_{opt} consists of updating $\nu(t)$ in the opposite direction of the gradient of E_{opt} . These results are also valid when a hyperbolic tangent activation function is used. Each iteration has two distinct stages. First, as described in Step 3, $\nu(t)$ is updated using the gradient of the term E_{opt} alone. Second, after each updating, $\nu(t)$ is projected directly in the valid subspace. This is an iterative process, in which $\nu(t)$ is first orthogonally projected in the valid subspace (7) and then thresholded so that its elements lie in the range $\nu(t) \in [L_{inf}, L_{sup}]$. Thus, the mapping of a linear programming problem using a modified Hopfield network consists of determining the weight matrices T_{conf} and T_{opt} , and the vectors I_{conf} and I_{opt} .

3. Transportation model

Transportation model is a special type of networks problems that for shipping a commodity from source (e.g., factories) to destinations (e.g., warehouse). Transportation model deal with get the minimum-cost plan to transport a commodity from a number of sources m to number of destination n . Let s_{ti} is the number of supply units required at source

$i = 1, 2, 3, \dots, m$ at $t = 1, 2, 3, \dots, T$, d_{tj} is the number of demand units required at destination $j = 1, 2, 3, \dots, n$ at $t = 1, 2, 3, \dots, T$ and c_{ij} represent the unit transportation cost for transporting the units from sources i to destination j . Using linear programming method to solve transportation problem, we determine the value of objective function which minimize the cost for transporting and also determine the number of unit can be transported from source i to destination j . If x_{tij} is number of units shipped from source i to destination j at t . The objective function

$$\begin{cases} \text{Minimize } Z_t = \sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{tij}, \\ \sum_{j=1}^m x_{tij} = s_{ti}, \text{ for } i = 1, 2, \dots, m, \\ \sum_{i=1}^n x_{tij} = d_{tj}, \text{ for } j = 1, 2, \dots, n, \\ x_{tij} \geq 0, \forall i, j. \end{cases}$$

A transportation problem is said to be balanced if the total supply from all sources equals the total demand in all destinations

$$\sum_{i=1}^m s_{ti} = \sum_{j=1}^n d_{tj},$$

otherwise it is called unbalanced. We assume that s_{ti} and d_{tj} are stochastic where $s_{ti} \propto W(\bar{s}_{ti}, \sigma_{si})$ and $d_{tj} \propto W(\bar{d}_{tj}, \sigma_{dj})$. This problem we can solve using methods of a linear programming problems by the modified Hopfield network.

4. Formulation of Linear Programming Problems by the Modified Hopfield Network

A linear programming problem is a problem of minimizing or maximizing a linear function in the presence of linear constraints of the inequality and/or equality type. Since equality constraints can be easily converted in inequality constraints, it is considered in this formulation only inequality constraints. Consider the following linear programming problem, with m -constraints and n variables, given by the following equations:

$$\begin{cases} \text{Minimize } E_{opt} = c^T \nu(t), \\ E_{conf}(\nu) : A^T \nu(t) \leq b, \\ \nu_{min} \leq \nu(t) \leq \nu_{max}. \end{cases} \quad (3)$$

where $A \in R^{n \times m}$, $b \in R^m$, and $c, \nu, \nu_{min}, \nu_{max} \in R^n$. The conditions in (3) define a bounded convex polyhedron. The vector ν must remain within this

polyhedron if it is to represent a valid solution for the optimization problem (3). A solution can be obtained by a modified Hopfield network, whose valid subspace guarantees the satisfaction of the condition (3). Moreover, the initial hypercube represented by the inequality constraints in (3) is mapped by the symmetric ramp function (3) used as a network activation function.

The parameters T_{conf} and I_{conf} are calculated by transforming the inequality constraints in (3) into equality constraints by introducing a slack variable $w \in R^n$ for each inequality constraint:

$$g_i(\nu(t)) = \sum_{j=1}^q \delta_{ij} w_j = 0,$$

where w_j are slack variables, treated as the variables ν_i , and δ_{ij} is defined by the Kronecker impulse function.

After this transformation the problem defined by equation (3) we can rewritten as:

$$\begin{cases} \text{Minimize } E_{opt} = c^T \nu(t), \\ E_{conf}(\nu) : \tilde{A}^T \tilde{\nu}(t) = \tilde{b}, \\ \nu_{min} \leq \tilde{\nu}(t) \leq \nu_{max}, i = 1, 2, \dots, n, \\ 0 \leq \tilde{\nu}(t) \leq \nu_{max}, i = n + 1, n + 2, \dots, N, \end{cases} \quad (4)$$

where $N = n + m$, and $\tilde{\nu}^T = \{\tilde{\nu}^T w^T\} \in R^N$ is a vector of extended variables. Note that E_{opt} does not depend on the slack variable w . If the rows of \tilde{A} are linearly independent, a solution for (4) is given by:

$$\tilde{\nu} = \tilde{A}(\tilde{A}^T \tilde{A})^{-1} \tilde{b}$$

and the expression of the valid subspace in (2) must take account this solution, i.e.,

$$I_{conf} = \tilde{A}(\tilde{A}^T \tilde{A})^{-1} \tilde{b}. \quad (5)$$

From (5) the parameter T_{conf} is derived as follows:

$$\begin{cases} \tilde{\nu} = T_{conf} \tilde{\nu} + I_{conf}, \\ \tilde{\nu} = T_{conf} \tilde{\nu} + \tilde{A}(\tilde{A}^T \tilde{A})^{-1} \tilde{b}. \end{cases} \quad (6)$$

Inserting the value of (4) in (6), the expression for Tconf is given by:

$$I_{conf} = I - \tilde{A}(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T,$$

where I is identity matrix. The parameters T_{opt} and I_{opt} in this case are such that the vector $\tilde{\nu}$ is updated in the opposite gradient direction that of the energy function E_{opt} . Since conditions (3) define a bounded convex polyhedron, the objective function (4) has a unique global minimum $T_{opt} = 0$. Thus, the parameters T_{opt} and I_{opt} are obtained by comparing equation (1) and (4). After this comparison, these parameters are given by:

$$I_{opt} = -c, I_{opt} = 0.$$

The architecture analysis of an algorithm based on neural networks allows to estimate the resources for network implementation, such as the number of necessary logical gates and/or amount of necessary memory.

5. Conclusions

In this paper, it is developed a modified Hopfield network for solving linear programming problems related to operations research. The internal parameters of the network were explicitly computed using the valid-subspace technique that guarantees the network convergence. Simulation results proved that the proposed network is a feasible alternative for solving such problems efficiently. A complexity analysis is also presented and shows the computational behavior involved in each neural approach used normally for solving linear programming problem. The analysis showed that this approach is effective for solving a wide range of technical and technological problems of design, construction and operation of transport systems. In this approach manages to create an efficient algorithm for minimizing the cost of design, construction and operation of such systems. The numerical analysis shows the effectiveness of the proposed approach.

Acknowledgments

The reported study was funded within the Agreement No 02.a03.21.0008 dated 24.11.2016 between the Ministry of Education and Science of the Russian Federation and RUDN University

References

1. *Hanialhossein Abolfazli, Seyed Masoud Asadzadeh, Salman Nazari-Shirkouhi, Seyed Mohammad Asadzadeh, Kamran Rezaie* Acta Polytechnica Hungarica. Vol. 11, No. 2, pp. 203 – 214 (2014)
2. *Calvert B.D.* Neural networks with an infinite number of cells, Journal of Differential Equations. Vol. 186, Issue 1, pp. 31 – 51 (2002)
3. *Calvert B.D., Zemanian A.H.* Operating points in infinite nonlinear networks approximated by finite networks, Trans. Amer. Math. Soc. Vol. 352, No 2 , 753 – 780, (2000).

4. *Henry D.* Geometric theory of semilinear parabolic equations. Lecture Notes in Mathematics. Springer-Verlag, Berlin (1981)
5. *Hopfield J.J.* Neurons with graded response have collective computational properties like those of two state neurons, Proc. Natl. Acad. Sci. USA. Vol. 81, pp. 3088 – 3092 (1984)
6. *Haiying Huang, Qiaosheng Du, Xibing Kang.* Global exponential stability of neutral high-order stochastic Hopfield neural networks with Markovian jump parameters and mixed time delays. ISA Transactions. Vol. 52, Issue 6, pp. 759 – 767 (2013)
7. *Ivan Nunes Da Silva, Andre Nunes De Souza, Jose Alfredo C. Ulson.* Neural Networks Applied in Linear Programming Problems: Design and Complexity Analysis. In: N. Mastorakis; V. Mladenov; B. Suter; L. J. Wang. (Org.). Advances in Scientific Computing, Computational Intelligence and Applications. 1 ed. Danvers-MA, USA: WSES Press, pp. 444–449 (2001)
8. *Rahman S.A., Ansari M.S., Moinuddin A.A.* Solution of linear programming problems using a neural network with nonlinear feedback. Radioengineering, 2012, vol. 21, no. 4, pp. 1171 – 1177.
9. *Seleznev B.E., Aleshin V.V., Pryalov S.N.* Matematicheskoe modelirovanie truboprovodnyh setej i sistem kanalov. Metody, modeli, algoritmy. — M., MAKSPRESS (2007)
10. *Sustekova D., Knutelska M.* Artificial intelligence applications especially the neural networks use in the road transport. Proceedings of the 13-th International Conference Reliability and Statistics in Transportation and Communication (RelStat 13), 16 — 19 October 2013, Riga, Latvia, pp. 265 – 271 (2013)
11. *Xiao Liang, Linshan Wang, Yangfan Wang, Ruili Wang.* Dynamical Behavior of Delayed Reaction-Diffusion Hopfield Neural Networks Driven by Infinite Dimensional Wiener Processes. IEEE Transactions on Neural Networks. Vol. 27, No 9, pp. 1816 – 1826 (2016)

UDC 004.75

Efficiency of Redundant Multipath Transmission of Requests Through the Network to Destination Servers

V. A. Bogatyrev, S. A. Parshutina

*Department of Software Engineering and Computer Systems
ITMO University
Kronverksky Pr. 49, St. Petersburg, 197101, Russia*

Abstract. It is not uncommon that delay-sensitive requests cannot be processed repeatedly in case of delivery failures, especially in real-time systems. Hence, there is a strong need to enhance reliability of requests' delivery to destination nodes (servers). It is therefore proposed to use redundant transmission of requests, when copies of a request are distributed concurrently through the network over multiple routes to a number of similar servers. However, the resulting increase in the initial flow of requests leads to the rise of the network load and the average residence time and possibly to the excess of the ultimate residence time. In the research, the usefulness of redundant distribution of requests through the network was estimated, with the maximized probability of successful delivery and the minimized average residence time. Thus, the range of efficiency of redundant multipath transmissions was defined. It was also found that there existed the optimal redundancy order for a given set of parameters, namely the intensity of the flow of incoming requests and the bit error rate.

Keywords: reliability, request, redundancy, multipath routing, queueing systems.

1. Introduction

The constantly growing complexity of distributed computing systems, which is ahead of the pace of the increasing reliability of storage, processing, and transmission devices, gives rise to the need for developing methods to ensure high reliability, availability, and security of distributed resources for data processing and storage [1–8]. High availability of distributed resources is particularly important for real-time systems, which are critical to delay-sensitive requests. It is commonly provided by introducing redundant elements into the structure of a given network and by designing the network in the way that lets arrange redundant data transmission and processing. In delay-sensitive systems, sending requests twice or more times to the destination, where they are somehow processed, is often impossible or pointless. This is due to rapid changes in the state of such systems, occurring under complex internal or external influence. As a result, methods to introduce redundancy into such systems turn out to be topical and prospective for in-depth exploration.

The proposed way of redundant data transmission should be employed when it is not possible to transmit data (a request, a data packet) through the network and/or process it at the destination node (a server) more than

once. This approach relies on the mechanism of multipath data transmission, particularly on multipath routing, which implies at least two appropriate routes (paths) between the sender and the receiver, or receivers — in case of multicast transmission. In this paper, the receiver is one of the servers belonging to the same group (or cluster). The servers have similar principal characteristics and serve incoming requests in the similar way.

Multipath routing is widely used for multicast transmission, when data packets undergo fragmentation and the resulting fragments are sent over different paths to be brought together again at the destination. This is intended to hasten the process of transmission of large data packets and to secure them, especially when it comes to the secret data. Packet fragmentation based on multipath routing is mostly applied to load balancing.

Alternatively, using several paths for data transmission proves beneficial for enhancing reliability and fault-tolerance of the system. The traditional approach thereto suggests that one route is regarded as the main, the "best" path from some point of view, i. e. a given criterion, or metrics: bandwidth, average residence time, router and link load, and so on. Other routes are considered as alternate ones, used in case the main path fails.

Another approach is presented in [9], where requests – more precisely, numerous copies of each request – are sent over multiple paths concurrently. There is no need to choose the "best" path, because all of them are regarded to be appropriate, based on some metrics, with no ranking. On the one hand, redundant distribution of requests through the network causes an increased network load and a higher average residence time as well as possible excess of the ultimate residence time. On the other hand, sending copies of a request over multiple paths results in the enhanced chance that at least one copy will be delivered successfully. The goal of current research is to estimate the efficiency of redundant distribution of requests through the network when their repeated transmission is impossible, and therefore, to discover if there is a way to resolve this technical contradiction.

2. Problem Statement

Current study focuses on the process of redundant distribution of requests through a network with the rigid requirements for the maximum period of time during which the requests are to be processed. In the known systems without redundancy, a request is sent from the source of requests to the single receiver, which might be a switching node or a server. In case of redundant systems, two or more copies of a request are transmitted from the sender to the multiple receivers (in this paper – servers) over a number of routes, as shown in Figure 1.

2.1. System Design

It is assumed that there are n appropriate paths to deliver a request to one of the servers; in this research, all those paths are considered disjoint.

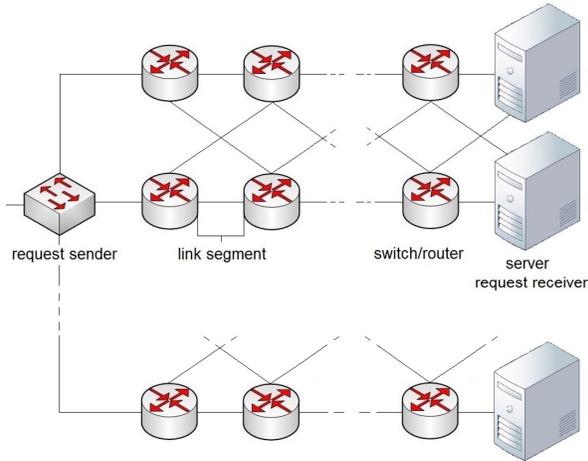


Figure 1. A possible network configuration for redundant data transmission.

Each path i leading from the source of requests to one of destination servers includes a set of d_i switches or routers K_{ij} and $d_i + 1$ link segments j between them; all units of network equipment have specified characteristics of reliability and performance. Each node K_{ij} can be represented as a single-channel non-preemptive M/M/1 queueing system with the infinite queue [10, 11].

Any errors and faults of switching nodes and link segments as well as of destination servers lead to the situation when requests cannot be delivered and served. Suppose that the bit error rate B_{ij} , which is a measure of transmission errors for the link segment j of the path i , is known. It is possible that a request reaches one of the destination servers, with no bit errors, but finds it unavailable due to faults or temporary shutdown or due to the fact that the server is overloaded with the requests which have arrived earlier than the one in question and been waiting for service in the queue of the server. Thus, the probability of a server's availability P_0 , introduced in [9], is taken into account.

Requests are distributed through the network over k ($k=2, \dots, n$) routes, where k is the rate of redundancy, or the redundancy rate. It is said that a request is delivered successfully if at least one copy of it arrives at the destination, contains no bit errors, and is placed into the server queue. Here, we admit that no copies of the request are discarded after one of them has been delivered successfully. This leads to the upper (pessimistic)

estimation of the average residence time, or the *time in the system*, needed for error-free transmission of a request from the source to the destination.

2.2. Probability of Error-free Transmission

To start with, let us estimate distribution of requests through the network on the basis of server availability P_0 and the bit error rate B_{ij} , which is in fact the probability of bit errors which might occur while transmitting a request over the link segment j of the path i .

The probability of successful transmission of a data packet over the path i can be calculated as following:

$$R_i = \prod_{j=1}^{d_i} (1 - B_{ij})^N,$$

where d_i is the number of link segments of the path i and N is the average data packet length in bits. Provided that all routes under consideration are identical, the bit error rate can be regarded as a constant and the formula (2.2) can be replaced with

$$R = (1 - B)^{Nd}.$$

When requests are being transmitted over k redundant paths concurrently, the probability of error-free transmission of at least one copy of some request over any path is

$$P_c = 1 - (1 - P_0 R)^k.$$

Here, P_0 is the probability of a server's being available, or ready to serve incoming requests, meaning that it is powered up and in working order and there is enough place in its queue for the requests coming from the network. In this paper, we suggest that all servers of the interest, i. e. belonging to the same group (cluster) and hence serving requests from the same source, have similar availability P_0 .

2.3. Average Residence Time

Let us estimate the average residence time for data packets being transmitted through the network to the destination, depending on the intensity of the request flow Λ . The delay T_i in the path i is defined as the sum of delays in all segments of that path:

$$T_i = \sum_{j=1}^{d_i} T_{ij},$$

where T_{ij} is the delay in the segment j of the path i , or the average time needed to transmit a data packet through the node j of the path i (with respect to transmission velocity of the output port):

$$T_{ij} = \sum_{j=1}^{d_i} \frac{v_{ij}}{1 - \frac{v_{ij}k\Lambda}{n}}.$$

Here, v_{ij} is the average time of transmission in the segment j of the path i and computed as follows:

$$v_{ij} = \frac{N}{L_{ij}},$$

where N is the average data packet length and L_{ij} is the transmission velocity.

When a data packet is sent over k identical routes simultaneously, the average residence time in the system T is defined as equal to the delay in some path – if the packet has been transmitted with no bit errors and accepted for further processing by the destination server. Assessing the average residence time for that path in the described way gives an upper (conservative) estimate. This is because the estimation does not take into account the probability that in at least one path used to transmit copies of data packets, the delivery time of the copy under consideration can be less than the resulting value of T , due to the stochastic nature of the processes of servicing.

2.4. Criteria of Efficient Delivery

Efficiency of redundant distribution of requests through the network depends on the average residence time T , which is to be minimized, and the probability of error-free transmission P_c , which is to be maximized. The combination of these two single criteria produces a new multiplicative criterion *Mult1*:

$$Mult1 = \frac{P_c}{T}.$$

For the situation when the ultimate residence time t_0 is known, another multiplicative criterion can be defined, namely

$$Mult2 = P_c(t_0 - T),$$

which describes the average stock time between the ultimate and the average residence time for the packets that have been delivered successfully.

Last but not least, let us estimate the efficiency of redundant transmission of requests through the network based on the following additive criterion, with the normalized residence time:

$$Add = P_c + \frac{t_0 - T}{t_0 - T_{min}},$$

where t_0 is the ultimate residence time, while T_{min} is the minimal residence time.

3. Calculations and Discussion

The efficiency of redundant distribution of requests (data packets) through the network to the destination servers is to be evaluated depending on the redundancy rate and according to the multiplicative criteria *Mult1* and *Mult2* and the additive criterion *Add*. Suppose that there are ten disjoint routes between the sender of requests and the receivers (servers), $n=10$. Each route contains $d=5$ intermediate nodes (switches or routers) and $d+1$ link segments between those nodes.

For the sake of simple calculations, we assume that the average time of transmission v_{ij} is the same for all segments and is equal to v . If the average data packet length N equals to 2048 (bit) and transmission velocity L_{ij} is 100 (Mbit/s), then the average time of transmission $v=2.148 \times 10^{-5}$ (s). Let us define the minimal residence time $T_{min}=dv=1.074 \times 10^{-4}$, considering that there are no queues in the switching nodes of any route, and the maximal residence time $t_0 = 7.6 \times 10^{-4}$.

Suppose that the probability of server availability P_0 is reasonably high: $P_0=0.9$; and the intensity of the flow of requests Λ varies, for example $\Lambda_1 = 9 \times 10^3$, $\Lambda_2 = 1 \times 10^4$, $\Lambda_3 = 2 \times 10^4$, $\Lambda_4 = 3 \times 10^4$ (1/s). The further estimations demonstrate that for the given values of other parameters, the plausible values of the bit error rate B are in the range 10^{-7} to 10^{-5} . The results of calculations, shown in Figure 2, are derived from the bit error rate $B=10^{-7}$.

Figure 2 illustrates the efficiency of redundant distribution of requests over k routes, which is computed on the basis of the multiplicative criterion *Mult1* (a), the multiplicative criterion *Mult2* (b), and the additive criterion *Add* (c). The displayed curves 1–4 in each of the three graphs correspond to the intensities of the request flow $\Lambda_1 - \Lambda_4$. It can be seen that sending data packets through the network in the redundant manner increases the efficiency of transmission but only up to a certain point (the maximum value of the criterion in question). This confirms the need to solve the problem of optimization, in order to determine the redundancy rate that provides the maximum efficiency across all proposed criteria. It is particularly important for the systems with the limited residence time.

It is important to note that the higher the intensity of the request flow Λ is, the lesser the values of all criteria under consideration are – *Mult1*, *Mult2*, and *Add*. With the increase of Λ , the maximum values of the criteria correspond to lesser values of the rate of redundancy k than before, as shown in Figure 2.

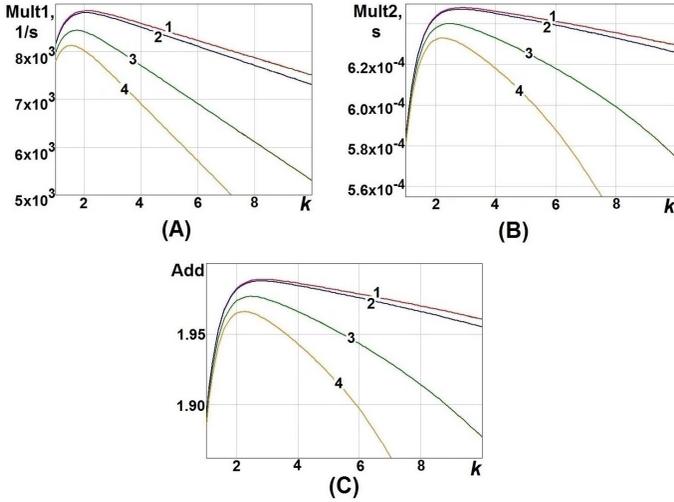


Figure 2. The efficiency of redundant data transmission estimated on the basis of criteria *Mult1(a)*, *Mult2(b)*, and *Add(c)*.

4. Conclusion

The results of the conducted study support the initial hypothesis that introducing redundancy in the process of data transmission in delay-sensitive systems, under conditions of possible transmission errors and server unavailability for processing incoming requests, is likely to enhance those systems' efficiency. The efficiency is understood in terms of the increased probability of errorless delivery of requests (data packets) and the decreased average residence time.

The range of efficiency of redundant distribution of requests through the network to the destination servers was determined.

On the basis of the proposed additive and multiplicative criteria, it was found that there existed the optimal redundancy order, or the optimal number of routes needed to transmit requests through the network in the redundant manner.

It was discovered that the redundancy order depended on the intensity of the flow of incoming requests as well as the probability of transmission errors (the bit error rate).

The results of the research may appear useful for those who need to decide for the redundancy order when sending data packets through the network, with the purpose to improve reliability of data transmission and accelerate distribution of requests to the destination servers.

Acknowledgments.

The study was fulfilled as a part of the research project "Methods for designing the key systems of information infrastructure" (State Registration Number is 615869) of ITMO University, St. Petersburg, Russia.

References

1. *Bogatyrev V.A.* Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer *Automatic Control and Computer Sciences*, vol. 33, No. 1, 1999, pp. 57-63.
2. *Bogatyrev V.A.* On interconnection control in redundancy of local network buses with limited availability *Engineering Simulation*, Vol. 16, No 4, 1999, pp. 463-469.
3. *Bogatyrev V.A.* An interval signal method of dynamic interrupt handling with load balancing *Automatic Control and Computer Sciences*, vol. 34, No. 6, 2000, pp. 51-57.
4. *Bogatyrev V.A.* Fault tolerance of clusters configurations with direct connection of storage devices *Automatic Control and Computer Sciences*, V. 45, No. 6, 2011, pp. 330—337.
5. *Bogatyrev V.A.* Exchange of Duplicated Computing Complexes in Fault tolerant Systems *Automatic Control and Computer Sciences*, Vol. 46, No. 5, 2011, pp. 268—276.
6. *Bogatyrev V.A., Bogatyrev S.V., Golubev I.Y.* Optimization and the Process of Task Distribution between Computer System Clusters *Automatic Control and Computer Sciences*, No. 3, 2012, pp. 103-111.
7. *Bogatyrev V.A., Bogatyrev A.V.* Functional Reliability of a Real-Time Redundant Computational Process in Cluster Architecture Systems *Automatic Control and Computer Sciences*, Vol. 49, No. 1, 2015, pp. 46-56.
8. *Arustamov S.A., Bogatyrev V.A., Polyakov V.I.* Back Up Data Transmission in Real-Time Duplicated Computer Systems. Proceedings of the First International Scientific Conference Intelligent Information Technologies for Industry (IITI-16), *Advances in Intelligent Systems and Computing*, Springer, 2016, Vol. 2, pp. 103-109.
9. *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*, 2016, Vol. 601, pp. 199-207.
10. *Aliev T. I.* The Synthesis of Service Discipline in Systems with Limits // *Distributed Computer and Communication Networks*. 18th International Conference, DCCN 2015. – Moscow, Russia, October 19—22, 2015. Revised Selected Papers, pp. 151-156.
11. *Aliev T.I., Rebezova M.I., Russ A.A.* Statistical Methods for Monitoring Travel Agencies *Automatic Control and Computer Sciences*, Vol. 49, No. 6, 2015, pp. 321-327.

UDC 517.937, 517.928.2, 519.217.2

Systems of Differential Equations of Infinite Order with Small Parameter and Countable Markov Chains

G. O. Bolotova*, S. A. Vasilyev*, D. N. Udin†

* *Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

† *IBM Osterreich Internationale Gesellschaft m.b.H.
Obere Donaustrasse 95, Wien, A-1020, , Austria*

Abstract. Tikhonov-type Cauchy problems are investigated for systems of ordinary differential equations of infinite order with a small parameter μ and initial conditions. It is studying the singular perturbed systems of ordinary differential equations of infinite order of Tikhonov-type $\dot{x} = f(x(t, g_x), y(t, g_y), t)$, $\mu \dot{y} = F(x(t, g_x), y(t, g_y), t)$ with the initial conditions $x(t_0, g_x) = g_x$, $y(t_0, g_y) = g_y$, where $x, f \in X$, $X \in \mathbf{R}^n$ are n -dimensional functions; $y, F \in Y$, $Y \subset l_1$ are infinite-dimensional functions and $t \in [t_0, t_1]$ ($t_0 < t_1 \leq \infty$), $t \in T$, $T \in \mathbf{R}$; $g_x \in X$ and $g_y \in Y$ are given vectors, $\mu > 0$ is a small real parameter. The results may be applied to the queuing networks, which arise from the modern telecommunications.

Keywords: systems of differential equations of infinite order, singular perturbed systems of differential equations, small parameter, countable Markov chains.

1. Introduction

The recent research of service networks with complex routing discipline in [16], [17], [18] transport networks [1], [4], [5] faced with the problem of proving the global convergence of the solutions of certain infinite systems of ordinary differential equations to a time-independent solution. Scattered results of these studies, however, allow a common approach to their justification. This approach will be expounded here. In work [11] the countable systems of differential equations with bounded Jacobi operators are studied and the sufficient conditions of global stability and global asymptotic stability are obtained. In [10] it was considered finite closed Jackson networks with N first come, first serve nodes and M customers. In the limit $M \rightarrow \infty$, $N \rightarrow \infty$, $M/N \rightarrow \lambda > 0$, it was got conditions when mean queue lengths are uniformly bounded and when there exists a node where the mean queue length tends to infinity under the above limit (condensation phenomena, traffic jams), in terms of the limit distribution of the relative utilizations of the nodes. It was derived asymptotics of the partition function and of correlation functions.

Cauchy problems for the systems of ordinary differential equations of infinite order was investigated A.N.Tihonov [13], K.P.Persidsky [12], O.A.Zhautykov [19], [20], Ju.Korobeinik [7] other researchers.

It was studied the singular perturbed systems of ordinary differential equations by A.N. Tihonov [14], A.B.Vasil'eva [15], S.A. Lomov [9] other researchers.

A particular our interest is the synthesis all these methods and its applications in telecommunications. In this paper we apply methods from [11] for the singular perturbed systems of ordinary differential equations of infinite order of Tikhonov-type.

2. Tikhonov-type Cauchy problems for systems of ordinary differential equations of infinite order with a small parameter

Let us consider Tikhonov-type Cauchy problems for systems of ordinary differential equations of infinite order with a small parameter μ and initial conditions:

$$\begin{aligned} \dot{x} &= f(x(t, g_x), y(t, g_y), t), \\ \mu \dot{y} &= F(x(t, g_x), y(t, g_y), t); \end{aligned} \quad (1)$$

$$\begin{aligned} x(t_0, g_x) &= g_x, \\ y(t_0, g_y) &= g_y, \end{aligned} \quad (2)$$

where $x, f \in X$, $X \in \mathbf{R}^n$ are n -dimensional functions; $y, F \in Y$, $Y \subset l_1$ are infinite-dimensional functions and $t \in [t_0, t_1]$ ($t_0 < t_1 \leq \infty$), $t \in T$, $T \in \mathbf{R}$; $g_x \in X$ and $g_y \in Y$ are given vectors, $\mu > 0$ is a small real parameter; $x(t, g_x)$ and $y(t, g_y)$ are solutions of (1)-(2). Given functions $f(x(t, g_x), y(t, g_y), t)$ and $F(x(t, g_x), y(t, g_y), t)$ are continuous functions for all variables. Let S is an integral manifold of the system (1)-(2) in $X \times Y \times T$. If any point $t^* \in [t_0, t_1]$ ($x(t^*), y(t^*), t^*$) $\in S$ of trajectory of this system has at least one common point on S this trajectory $(x(t, G), y(t, g), t) \in S$ belongs the integral manifold S totally. If we assume in (1)-(2) that $\mu = 0$ than we have a degenerate system of the ordinary differential equations and a problem of singular perturbations

$$\begin{aligned} \dot{x} &= f(x(t, g_x), y(t), t), \\ 0 &= F(x(t, g_x), y(t), t); \\ x(t_0, g_x) &= g_x, \end{aligned} \quad (3)$$

where the dimension of this system is less than the dimension of the system (1)-(2), since the relations $F(x(t), y(t), t) = 0$ in the system (3) are the algebraic equations (not differential equations). Thus for the system (3) we can use limited number of the initial conditions then for system (1)-(2). Most natural for this case we can use the initial conditions $x(t_0, g_x) = g_x$ for the system (3) and the initial conditions $y(t_0, g_y) = g_y$ disregard otherwise we get the overdefined system. We can solve the system (3) if

the equation $F(x(t), y(t), t) = 0$ could be solved. If it is possible to solve we can find a finite set or countable set of the roots $y_q(t, g_x) = u_q(x(t, g_x), t)$ where $q \in \mathbf{N}$.

If the implicit function $F(x(t), y(t), t) = 0$ has not simple structure we must investigate the question about the choice of roots. Hence we can use the roots $y_q(t, g_x) = u_q(x(t, g_x), t)$ ($q \in \mathbf{N}$) in (3) and solve the degenerate system

$$\begin{aligned} \dot{x}_d &= f(x_d(t, g_x), u_q(x_d(t, g_x), t), t); \\ y_d(t_0, g_x) &= g_x. \end{aligned} \quad (4)$$

Since it is not assumed that the roots $y_q(t, g_x) = u_q(x(t, g_x), t)$ satisfy the initial conditions of the Cauchy problem (1)-(2) ($y_q(t_0) \neq g_x$, $q \in \mathbf{N}$), the solutions $y(t, g_y)$ (1)-(2) and $y_q(t, g_x)$ do not close to each other at the initial moments of time $t > 0$. Also there is a very interesting question about behaviors of the solutions $x(t, g_x)$ of the singular perturbed problem (1)-(2) and the solutions $x_d(t, g_x)$ of the degenerate problem (4). When $t = 0$ we have $x(t_0, g_x) = x_d(t_0, g_x)$. Do these solutions close to each other when $t \in (t_0, t_1]$? The answer to this question depends on using roots $y_q(t, g_x) = u_q(x(t, g_x), t)$ and the initial conditions which we apply for the systems (1)-(2) and (3).

3. Local existence theorem for Cauchy problems for systems of ordinary differential equations of infinite order

Let Tikhonov-type Cauchy problems for systems of ordinary differential equations of infinite order with a small parameter $\mu > 0$ and initial conditions (1)-(2) has a form:

$$\begin{aligned} \dot{z} &= P(z(t, G, \mu), t, \mu); \quad z(t_0, G, \mu) = G, \\ z &= (x_1, x_2, \dots, x_n, y_1, y_2, \dots)^T, \\ P(z(t, G, \mu), t, \mu) &= (f_1, f_2, \dots, f_n, \mu^{-1}F_1, \mu^{-1}F_2, \dots)^T, \\ G &= (g_{x1}, g_{x2}, \dots, g_{xn}, g_{y1}, g_{y2}, \dots)^T \end{aligned} \quad (5)$$

where $P(z(t, G, \mu), t, \mu)$ is the infinite-dimensional function; G is the given vector; $t \in [t_0, t_1]$ ($t_0 < t_1 \leq \infty$).

Let $z(t, G, \mu)$ be a continuously differentiable solution of the Cauchy problems (5) then there are $\Phi(t, G, \mu) = \partial z(t, G, \mu) / \partial G$, $\Psi(t, G, \mu) = \partial z(t, G, \mu) / \partial \mu$ where $\Phi(t, G, \mu)$ and $\Psi(t, G, \mu)$ satisfy of the system of ordinary differential equations in variations:

$$\dot{z} = P(z(t, G, \mu), t, \mu),$$

$$\begin{aligned}\dot{\Phi}(t, G, \mu) &= J_z(t, G, \mu)\Phi(t, G, \mu), \\ \dot{\Psi}(t, G, \mu) &= J_z(t, G, \mu)\Psi(t, G, \mu) + \Lambda_\mu(t, G, \mu); \quad (6) \\ z(t_0, G, \mu) &= G, \quad \Phi(t_0, G, \mu) = I, \quad \Psi(t_0, G, \mu) = 0, \quad t_0 \in T,\end{aligned}$$

where $J_z(t, G, \mu) = (\partial P_i / \partial z_j)_{i,j=1}^\infty$ is Jacobi matrix, I is an identity operator and $\Lambda_\mu(t, G, \mu) = (\partial P_i / \partial \mu)_{i=1}^\infty$ is a vector.

Theorem 1 (local existence theorem). *Let $P(z(t, G, \mu), t, \mu)$, $J_z(t, G, \mu)$, $\Lambda_\mu(t, G, \mu)$ be continuous and meet Gelder's local condition with $z \in U_\epsilon(G)$ then the system (6) has only one solution, which meet the conditions $z(t_0, G, \mu) = G$, $z(t, G, \mu) \in U_\epsilon(G)$. Thus $z(t, G, \mu)$ continuously differentiable with respect to the initial condition, and its derivative meet the equation (6).*

Proof. This statement is following from [3] (theorem 3.4.4) when the unlimited operator be $A = 0$. *End proof.*

The behavior of the solution $z(t, G, \mu)$ (5) and the nonnegative condition for the off-diagonal elements of the matrix $J_z(t, G, \mu)$ is demonstrated by the following theorem.

Theorem 2. *Let the solution z of (5) be $z(t, G, \mu) \in l_1$ for any $t \geq 0$, $G \in l_1$ and μ . The following claims are equal: (i) the off-diagonal elements $J_z(t, G, \mu)$ are non-negative for any G ; (ii) for any G and any vector $h \in l_1$, $h \geq 0$, $z(t, G + h, \mu) \geq z(t, G, \mu)$.*

Proof. Let us examine a convex set Z , and $z(t, G, \mu) \in Z$ for any $G \in Z$, derivative $\Phi(t, G, \mu)$ of function $z(t, G, \mu)$ can be specify by simultaneous equations (6). In that case the following formula is fair for any $G^0, G^1 \in Z$:

$$\begin{aligned}z(t, G^1, \mu) - z(t, G^0, \mu) &= \\ &= \int_0^1 \Phi(t, \gamma(s), \mu)(G^1 - G^0) ds\end{aligned} \quad (7)$$

where $\gamma(s) = (1 - s)G^0 + sG^1$, $0 \leq s \leq 1$.

In fact the function $z(t, G, \mu)$ transfer the segment $\gamma(s)$ into the curve $z(t, \gamma(s), \mu)$ in (7). The following formula is fair because of the continuous differentiability of function $z(t, G, \mu)$

$$z(t, \gamma(\tau), \mu) = z(t, G^0, \mu) + \int_0^\tau \frac{\partial z(t, \gamma(s), \mu)}{\partial s} ds.$$

By the formula of complex derivative

$$\frac{\partial z(t, \gamma(s), \mu)}{\partial s} = \frac{\partial z}{\partial G}(\gamma(s))\gamma'(s)$$

Recalling that $\partial z/\partial G = \Phi$ and $\gamma'(s) = G^1 - G^0$, with $\tau = 1$ we get (7). Let us suppose that statement (i) is fair. So because of (7)

$$z(t, G + h, \mu) - z(t, G, \mu) = \int_0^1 \Phi(t, \gamma(s), \mu) h ds$$

where $\gamma(s) = G + sh, 0 \leq s \leq 1$. Because of non-negativeness of function $J_z(t, G, \mu)$ outside of diagonal from (7) we get $\Phi(t, \gamma(s), \mu) \geq 0$, so $\Phi(t, \gamma(s), \mu)h \geq 0$ whence we get statement (ii).

Let us suppose that (ii) is fair. Under the conditions of Theorem 1 P, J_z with $z \in U_\epsilon(G)$ be continuous and meet Gelder's local condition. Let Gelder's local condition be $\|P\| < M_0, \|J\| < M_1$, and there are numbers $\delta > 0, \delta = \min(\epsilon/M_0, 1/M_1)$. Let $z(t, G, \mu) = G + z^*(t, G, \mu)$ be a solution of (7), where $z^*(t, G, \mu)$ is a fixed point of Picard's mapping $(\prod \theta)(t) = \int_{t_0}^t P(G + \theta(\tau)) d\tau$ under conditions $t \in [t_0 - \delta_1, t_0 + \delta_1], \delta_1 < \delta$. Mapping \prod is contraction with coefficient $\lambda = \delta_1 M_1 < 1$. Consider the approximation to solution $\tilde{z}(t, G, \mu) = G + \tilde{z}^*(t, G, \mu) = G + (t - t_0) P(z(t, G, \mu), t, \mu)$. We can see that

$$\begin{aligned} & \|\tilde{z}(t, G, \mu) - z(t, G, \mu)\| = \\ & = \|\tilde{z}^*(t, G, \mu) - z^*(t, G, \mu)\| \leq \\ & \leq \frac{1}{1 - \lambda} \|\prod \tilde{z}(t, G, \mu) - \tilde{z}(t, G, \mu)\|, \\ & \quad \prod \tilde{z}(t, G, \mu) - \tilde{z}(t, G, \mu) = \\ & = \int_{t_0}^t P(G + (\tau - t_0)P) d\tau - \int_{t_0}^t P d\tau = \\ & = \int_{t_0}^t (P(G + (\tau - t_0)P) - P) d\tau = D. \end{aligned}$$

Because of the derivative of the function P is limited and P meet Gelder's local condition with the constant M_1 , where $\|P(G + (\tau - t_0)P(G)) - P(G)\| \leq M_1 \|(\tau - t_0)P(G)\| \leq M_0 M_1 |\tau - t_0|$, so $\|D\| \leq M_0 M_1 (t - t_0)^2 / 2(1 - \lambda)$, or $\|\tilde{z}(t, G, \mu) - z(t, G, \mu)\| \leq M_0 M_1 (t - t_0)^2 / 2(1 - \lambda)$. Using this estimation and for all small $\zeta > 0$ we have that

$$0 \leq z(t, G + \zeta e_j, \mu) - z(t, G, \mu) = \zeta e_j + (t - t_0)[P(G + \zeta e_j) - P(G)] + \gamma(G, t),$$

where $\|\gamma(G, t)\| \leq M_0 M_1 (t - t_0)^2 / 2(1 - \lambda)$ and e_j is a vector, which has all coordinates equal to 0 but j -th coordinate equal to 1. Component $i \neq j$ of this inequality is given by $0 \leq (t - t_0)[P_i(G + \zeta e_j) - P_i(G)] + \gamma_i(G, t)$. Dividing by $t - t_0 > 0$ and directing $t \rightarrow t_0$ on the right, considering

$\gamma_i(G, t)/(t - t_0) \rightarrow 0$ we get $0 \leq P(G + \zeta e_j) - P(G)$. Let us divide last expression by ζ and direct $\zeta \rightarrow 0$

$$0 \leq \lim_{\zeta \rightarrow 0^+} \frac{P(G + \zeta e_j) - P(G)}{\zeta} = \frac{\partial P_i}{\partial G_i} = J_{ij}$$

what is mean the fairing of statement (i). *End proof.*

Theorem 3. *Let Φ be Markovian mapping and $G^0, G^1 \in X$, $t \geq 0$, $\mu > 0$ than $\|z(t, G^1, \mu) - z(t, G^0, \mu)\| \leq \|G^1 - G^0\|$.*

Proof. Using (6) from the proofing of theorem 4 we have

$$\|z(t, G^1, \mu) - z(t, G^0, \mu)\| \leq \int_0^1 \|\Phi(t, \gamma(s))(G^1 - G^0)\| ds$$

Let function $\Phi(t, \gamma(s))$ is Markovian mapping for any

$$\|t \geq 0, s \in [0, 1], \|\Phi(t, \gamma(s))(G^1 - G^0)\| \leq \|G^1 - G^0\|.$$

Estimating the integral, considering this inequality, we get required. *End proof.*

This theorem shows us the following sufficient condition for the boundedness of the norm-solution $z(t, G, \mu)$.

Corollary fact from theorem 3. *Let $\exists G^* \in X: z(t, G^*, \mu) = G^*$. Then $\|z(t, G, \mu) - G^*\| \leq \|G - G^*\|$ with $t \geq 0, G \in X$.*

This fact we can use for solutions analysis of the systems (5).

4. Conclusions

The boundaries of applications and possible generalizations. Some works in the routing disciplines. All systems can be analyzed for the global stability but with some condition that the convergence to the steady-state solution will not coordinate-wise, but the norm. We have seen that the most serious constraints of our methods are non-negativity of the Jacobi matrix off-diagonal elements and the availability of the first integral, which equal to the sum of the components. It would be interesting to understand the physical meaning of these conditions. It is necessary to remember that such systems describes the behavior of the queue lengths on the devices. Roughly speaking, z_k is the proportion of units in the queue for a service, to which there is at least k requests (including requests, which are serviced at the moment). Non-negative elements of the Jacobi matrix indicate that the rate of change of z_k (i.e., the time derivative of z_k) can only grow at the expense of z_j with $j \neq k$. It can be reduced (or decrease) only due to u_k . Thus, with the increase of the portion of queues with a minimum number of requests j in the system, the percentage change in intensity with the minimum number of queues requests $k \neq j$ can only increase.

Acknowledgments

The reported study was funded within the Agreement No 02.a03.21.0008 dated 24.11.2016 between the Ministry of Education and Science of the Russian Federation and RUDN University

References

1. *Afanassieva L.G., Fayolle G., Popov S. Yu.* Models for Transportation Networks. *J. Math. Science.* Vol.84, Issue 3, pp. 1092–1103 (1997)
2. *Daletsky Y.L., Krein M.G.* Stability of solutions of differential equations in Banach space. Moscow, Science Pub. (1970)
3. *Henry D.* Geometric theory of semilinear parabolic equations. Lecture Notes in Mathematics. Springer-Verlag, Berlin (1981)
4. *Khmelev D. V., Oseledets V.I.* Mean-field approximation for stochastic transportation network and stability of dynamical system: Preprint No. 434 of University of Bremen (1999)
5. *Khmelev D. V.* Limit theorems for nonsymmetric transportation networks. *Fundamentalnaya i Prikladnaya Matematika.* Vol. 7, no. 4, 1259 – 1266.(2001)
6. *Kirstein B. M., Franken D. E., Stoian D.* Comparability and monotonicity of Markov processes. *Theory of probability and its applications.* Vol. 22, Issue 1, pp.43–54 (1977)
7. *Korobeinik Ju.* Differential equations of infinite order and infinite systems of differential equations. *Izv. Akad. Nauk SSSR Ser. Mat.* Vol. 34, pp. 881 – 922 (1970)
8. *Krasnoselsky M.A., Zabreyko P.P.* Geometrical methods of nonlinear analysis. Springer-Verlag, Berlin (1984)
9. *Lomov S. A.* The construction of asymptotic solutions of certain problems with parameters. *Izv. Akad. Nauk SSSR Ser. Mat.* Vol. 32, pp. 884 – 913 (1968)
10. *Malyshev V. and Yakovlev A.* Condensation in large closed Jackson networks. *Ann. Appl. Probab.* Vol. 6, no. 1, pp. 92–115 (1996)
11. *Oseledets V. I., Khmelev D. V.* Global stability of infinite systems of nonlinear differential equations, and nonhomogeneous countable Markov chains. *Problemy Peredachi Informatsii (Russian),* Vol. 36 , Issue 1, pp. 60–76 (2000)
12. *Persidsky K.P.* *Izv. AN KazSSR, Ser. Mat. Mach.,* Issue 2, pp. 3–34. (1948)
13. *Tihonov A. N.* Uber unendliche Systeme von Differentialgleichungen. *Rec. Math.* Vol. 41, Issue 4, pp. 551–555 (1934).
14. *Tihonov A. N.* Systems of differential equations containing small parameters in the derivatives. *Mat. Sbornik N. S.* Vol. 31, Issue 73, pp. 575 – 586 (1952)
15. *Vasil'eva A. B.* Asymptotic behaviour of solutions of certain problems for ordinary non-linear differential equations with a small parameter

- multiplying the highest derivatives. *Uspehi Mat. Nauk.* Vol. 18, Issue 111, no. 3, 15 – 86 (1963)
16. *Vvedenskaya N.D., Dobrushin R.L., Kharpelevich F.I.* Queueing system with a choice of the lesser of two queues – the asymptotic approach. *Probl. inform.* Vol. 32, Issue 1, pp.15–27 (1996)
 17. *Vvedenskaya N.D., Suhov Yu.M.* Dobrushin’s Mean-Field Approximation for a Queue with Dynamic Routing. *Markov Processes and Related Fields.* Issue 3, pp. 493–526 (1997)
 18. *Vvedenskaya N.D.* A large queueing system with message transmission along several routes. *Problemy Peredachi Informatsii.* Vol. 34, no. 2, pp. 98–108 (1998)
 19. *Zhautykov O. A.* On a countable system of differential equations with variable parameters. *Mat. Sb. (N.S.).* Vol. 49, Issue 91, pp. 317 – 330 (1959)
 20. *Zhautykov O. A.* Extension of the Hamilton-Jacobi theorems to an infinite canonical system of equations. *Mat. Sb. (N.S.).* Vol. 53, Issue 95, pp. 313 – 328 (1961)

UDC 004.4

Accelerated regenerative simulation of degradation process in the system with gradual and sudden failures

A. B. Borodina^{*†}, D. V. Efrosinin[‡], E. V. Morozov^{*†}

** Institute of Applied Mathematical Research
Karelian Research Center RAS*

Pushkinskaya str. 11, Petrozavodsk, Karelia, 185910, Russia

*† Department of Mathematics and Information Technologies
Petrozavodsk State University*

Lenin Str. 33, Petrozavodsk, Karelia, 185910, Russia

*‡ Department of Applied Probability and Informatics
Peoples' Friendship University of Russia*

Miklukho-Maklaya str. 6, Moscow, 117198, Russia

Abstract. The optimal control of systems with gradual and sudden failures includes the evaluation problem of some stochastic parameters such as the failure probability, the average time before failure, the average cycle length etc. Some parameters may be calculated analytically if the degradation process is Markovian. But a problem arises for a non-Markov process. To this end, in more general cases we need to use simulation. If the failure within a regeneration cycle is a rare event then the regenerative version of Splitting method can be applied. We propose the Splitting method modification for estimation of degradation process characteristics. The point estimators are built by simulation of regenerative cycles.

Keywords: rare event simulation, splitting method, regenerative approach, degradation process.

1. Introduction

Consider a degradation process $\{X(t), t \geq 0\}$ in the system with gradual and sudden failures discussed in [1]. By simulation it is planned to construct the estimator for failure probability (which assumed to be rather small) and receive the other stochastic characteristics.

Following [1] we similarly use the regenerative approach but offer speed-up simulation technique (Splitting method) which can be particularly useful for non-Markov case. It is proposed to apply the regenerative version of the Splitting method [5, 6] to estimate the probability of sudden failure.

The suggested approach also allows to estimate the cycle length before refusal, length of cycle without failure, the total length of cycle and calculate the reliability function asymptotic.

2. Splitting for degradation process

Retaining the notation from [1] describe the degradation process in more details using Fig. 1), where the system state space defined by set

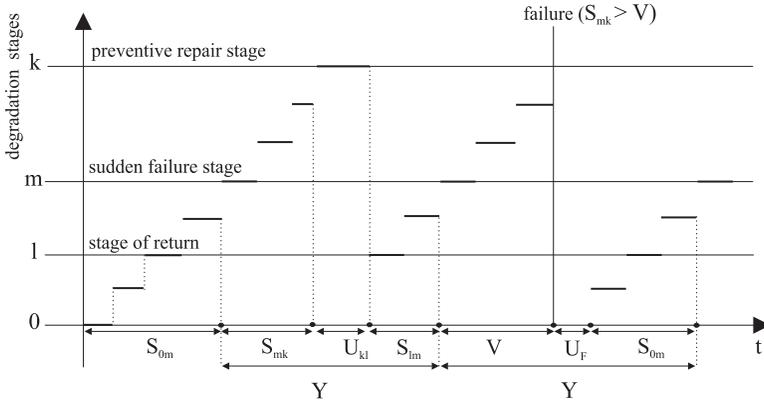


Figure 1. Realization of degradation managed process with two types of cycle

$E = \{0, 1, \dots, l, \dots, m, \dots, k, F\}$. The process $X(t)$ starts from 0 stage and *restored* after stage k . Denote by l the *return stage*, by m the stage of *sudden failure*, by T_i the time of sitting in the stage i until transition to the $j + 1$ stage, $S_{ij} = \sum_{p=i}^{j-1} T_p$. Since the degradation process can be accelerated in time $T_i, 0 \leq i \leq k$ are independent but not necessary identical distributed r.v. To enable analytical calculations in [1] T_i are taken exponentially distributed but with different parameters.

The figure Fig. 1) shows that after hitting m stage either sudden failure happens during the time V or the transition to preventive repair stage occurs during the time S_{mk} . So, there are two types of cycles in the system: with and without failure.

The main purpose is to estimate $p_F := P(S_{mk} > V)$ at the regeneration cycle and compare the results obtained by Splitting with Monte-Carlo results. A standard Splitting algorithm [2–4] accelerates the rare event emergence due to resplits of the process trajectory in defined levels. Features of the Splitting algorithm allow to give the regenerative interpretation of the method [5, 6].

In our setting, we focus on modification of Splitting where the stages of degradation process used as a thresholds with R_i number of copies of r.v. $T_i, m < i \leq k - 1, R_m = 1$. Splitting procedure begins at the regeneration moment, when the process reaches the stage m . As in a standard Splitting, the total number of paths is $D = R_m \cdot \dots \cdot R_{k-1}$ and each path can be treated as a part of regeneration cycle, started from stage m and truncated at the stage k .

In this way each path started from m gives the *bunch* of D dependable regeneration cycles. The cycles from different bunches are independent by construction. The cycles from one bunch have a common parts before

stage m and after stage k . The total number of bunches is the random variable R_{m-1} and is related to the number of transitions from stage $m-1$ to m . So, the p_F estimator is

$$\hat{p}_F = \frac{N_F}{R_{m-1} \cdots R_{k-1}},$$

where N_F is the number of cycles with failure.

3. Conclusions

In this paper, we propose the accelerated technique for rare event simulation for degradation process. This approach complements the analytical methods discussed in [1] and can be useful when the system parameters are not exponential. The simulation results show a significant advantage of the proposed method in comparison with standard Monte-Carlo simulation.

Acknowledgments

The work is supported by the RFBR, projects 15-07-02341, 15-07-02354, 15-07-02360, 16-37-60072 and by the Program of strategy development of Petrozavodsk State University.

References

1. *Efrosinin D. V., Farhadov M. P.* Optimal management of the system with the gradual and sudden failures. Dependability. 2009. No. 1 (28). PP. 27-42. (in Russian)
2. *Garvels M.* PhD Thesis. "The splitting method in rare event simulation", The University of Twente, The Netherlands May, 2000.
3. *Glasserman P., Heidelberger P., Shahabuddin P., and Zajic T.* Splitting for rare event simulation: analysis of simple cases. Proceedings of the 1996 Winter Simulation Conference, 1996, 302-308.
4. *Heidelberger P.* Fast simulation of rare events in queuing and reliability models, Performance Evaluation of Computers and Communications Systems Springer-Verlag, LN in Computer Sci., v. 729, 1993, 165-202.
5. *Borodina A., Morozov E.* Accelerated consistent estimation of a high load probability in M/G/1 and GI/G/1 queues. Journal of Mathematical Sciences, July 2014, Vol. 200, Issue 4, pp. 401-410.
6. *Borodina A. V.* PhD Thesis. Regenerative modification of the splitting method for estimating the overload probability in queuing systems, Petrozavodsk State University, 2008. (in Russian)

УДК 621.395

Анализ структуры сетей связи на базе беспилотных летательных аппаратов

Ч. Н. Дао, А. И. Парамонов

Кафедра Сетей связи и передачи данных, Санкт-Петербургский государственный университет телекоммуникаций им. проф.

М.А.Бонч-Бруевича,

пр.Большевиков д.22, корп.1, Санкт-Петербург, Россия, 193232

Аннотация. В статье приводятся результаты анализа способов построения сетей связи на базе беспилотных летательных аппаратов (БПЛА). Рассмотрены основные варианты структур сетей связи образуемой группой БПЛА, как между БПЛА в группе, так и между БПЛА и центром управления. Рассмотрены основные отличительные особенности и характеристики различных способов организации сети, их достоинства и недостатки. Полученные в работе результаты дают возможность выбирать оптимальный способ организации сети БПЛА в зависимости от требований, определяемых целевым назначением сети и ее параметрами, такими как количество БПЛА, число групп, масштабы сети, характеристики движения.

Ключевые слова: Беспилотные летающие аппараты, Централизованные связи, Децентрализованная сеть.

1. Введение

В настоящее время исследования в области всепроникающей сенсорной сети постепенно переходят от изучения характеристик на плоскости к моделям в трехмерном пространстве – это летающие сенсорные сети [1]. При этом в качестве летающих узлов используются общедоступные беспилотные летающие аппараты (БПЛА). БПЛА широко используются как в области задач военного характера, так и для решения задач гражданского назначения (разведчики, мишени, экология сельское хозяйство, управление транспортными потоками на автотрассах и т.д.) Использование квадрокоптеров или иных БПЛА для развертывания сенсорной сети, дает ряд преимуществ, так как они способны быстро перемещаться, не требуют наличия автодорог, а также могут выступать в качестве мобильных шлюзов, с другими сетями связи, в том числе и с сетью связи общего пользования. Использование БПЛА во многих случаях позволяет существенно снизить стоимость выполнения различных задач. Широкая доступность и низкие цены на БПЛА расширяют круг их применения. Задача покрытия - это ключевой вопрос для организации связи между БПЛА. Для достижения этой цели требуется решать задачи такие задачи как:

-задачи управления БПЛА (связь с центром управления или другими БПЛА, в зависимости от способа организации управления),

-задачи получения информации (сенсоры, видео, и.т.д.) и передачи ее в центр управления,

-задача группового управления БПЛА очень перспективна, так как применение группы БПЛА существенно повышает эффективность выполнения задания, при этом многие задачи существенно упрощаются. Появляется возможность проведения мониторинга больших территорий за короткое время без упущения важных деталей [2].

В связи с этим представляет интерес задача обеспечения и поддержания эффективной связи между БПЛА. В общем случае БПЛА могут двигаться с разной скоростью и по различным траекториям, что представляет очевидные сложности обеспечения непрерывной связи между ними. Сеть БПЛА имеет фиксированное или переменное число узлов и связей, а также с изменяющиеся координаты узлов. Поэтому, обеспечение связи между узлами такой сети требует выбора подходящей сетевой архитектуры.

2. Основная часть

Рассмотрим две основных архитектуры построения сети связи для группы БПЛА такие как централизованная и децентрализованная сети [3] и их комбинации. На рис.1. схематично изображены архитектуры сетей для централизованного, децентрализованного и комбинированных способов.

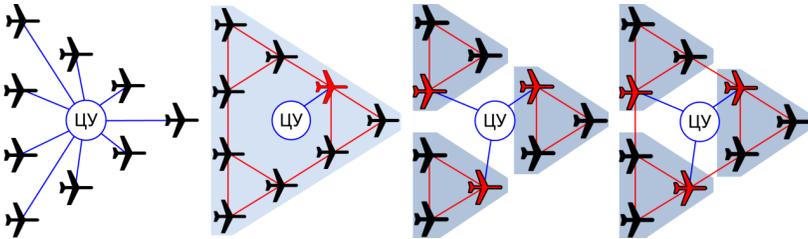


Рис. 1. Способы организации сети БПЛА, слева направо: централизованный, децентрализованный (ad hoc сеть), многогрупповой и многослойный способы

Одним из основных показателей функционирования сети связи с БПЛА является связность, показателем которой является вероятность связности p_{con} . При работе группы БПЛА можно различать связность БПЛА с центром управления p_c и связность между БПЛА p_u .

Централизованная сеть. В этом случае все БПЛА подключены непосредственно к одному или более центральных узлов (наземные станции) и все связи между БПЛА организуются маршрутами через наземную станцию (сеть станций). В такой структуре область связи

с БПЛА ограничена возможностями радиоканала БПЛА – наземная станция, также весь трафик создаваемой группой устройств должен обслуживаться наземной станцией, что может привести к перегрузке и как следствие к потере связи со всеми БПЛА. При необходимости связи между БПЛА трафик также должен обслуживаться наземной станцией, что приводит к росту задержки. При отказе наземной станции или связи с ней, вся сеть БПЛА полностью теряет связность. Эта архитектура достаточно часто используется в различных приложениях, когда не требуется организации связи между БПЛА в группе. В данной архитектуре сети требуется обеспечение достаточной пропускной способности линий связи БПЛА - наземная станция, достижимая величина которой ограничивает область действия всей группы. В данном случае вероятность связности между i -м БПЛА и центром управления p_c определяется только вероятностью существования канала связи между БПЛА и центром управления p_{uc} , т.е.

$$p_c = p_{uc}$$

Вероятность связности между двумя БПЛА i и j определяется как вероятность существования двух каналов

$$p_u = p_{ic}p_{cj}$$

где p_{ic} – вероятность существования канала между i -м БПЛА и центром управления; p_{cj} – вероятность существования канала между центром управления и j -м БПЛА.

При $p_{ic} = p_{cj} = p_{uc}$

$$p_u = p_{uc}^2$$

Децентрализованная сеть. В большинстве случаев, БПЛА системы работают в динамично изменяющейся среде, реализация алгоритма группового взаимодействия БПЛА использует схему связи, в которой один из группы БПЛА является «головным». В случае изменения условий распространения радиосигнала, некоторые из БПЛА могут потерять связность. Если сеть БПЛА поддерживает архитектуру FANET, то связь с головным узлом может быть обеспечена через другие БПЛА. Децентрализованная сеть является наиболее автономной, она обеспечивает надежную связь между БПЛА т.к. не имеет общей точки, потеря связности с которой приводит к отказу сети. Такая сеть обеспечивает функционирование при отказах отдельных связей и узлов. Такая система связи требует изменений на уровнях MAC и сетевом уровне, которые должны обеспечивать возможность самоорганизации, быть толерантными к задержкам, иметь гибкий автоматизированный контроль. Данная архитектура организации сети обеспечивает наибольшую надежность. Существуют три архитектуры такой сети: БПЛА ad-нос-сеть, многогрупповая БПЛА ad-нос-сеть и много-слойная БПЛА ad-нос-сеть.

ad-нос-сеть БПЛА. В БПЛА ad-нос-сети выделяется головной БПЛА, который служит шлюзом ad-нос сети. Через него проходят все

маршруты передачи данных между наземной станцией и другими БПЛА, как это показано на рис.1. Приемопередающее устройство БПЛА может быть недорогим и иметь малую массу. Данный способ позволяет расширить область действия БПЛА за счет организации связей внутри группы. Вероятность связности между i -м БПЛА и центром управления определяется вероятностью существования канала связи p_{ic} , а также вероятностью существования маршрута через ad hoc сеть p_a и вероятностью существования канала хотя бы одного БПЛА с центром управления. Структура маршрута БПЛА - центр управления в ad hoc сети приведена на рис.2.

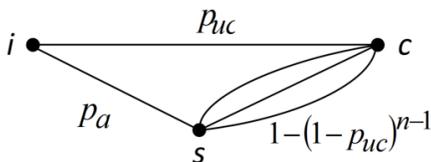


Рис. 2. Структура маршрута БПЛА - центр управления в ad hoc сети

$$p_c = 1 - (1 - p_{uc})(1 - p_a(1 - (1 - p_{uc})^{n-1}))$$

где p_a – вероятность связности между БПЛА в группе;
 p_{uc} – вероятность связности БПЛА с центром управления.
 n – количество БПЛА в группе.

Вероятность связности между двумя БПЛА i и j определяется вероятностью существования маршрута через ad hoc сеть p_a и вероятностью связи через центр управления (рис.3).

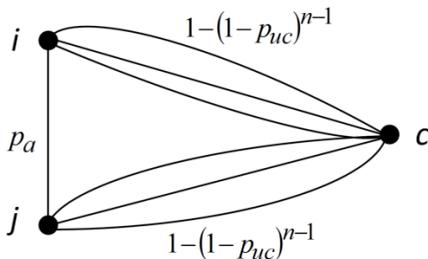


Рис. 3. Структура маршрута БПЛА - БПЛА в ad hoc сети

$$p_u = 1 - (1 - p_a)(1 - (1 - (1 - p_{uc})^{n-1})^2)$$

где p_a – вероятность связности между БПЛА в группе;
 p_{uc} – вероятность связности БПЛА с центром управления;
 n – количество БПЛА в группе.

Многогрупповая ad-нос-сеть БПЛА. Эта архитектура схожа с ad-нос-сетью БПЛА. В данной архитектуре формируются несколько групп БПЛА, в каждой из которых выделяется головной БПЛА, который соединяется с наземной станцией, как это показано на рис. 3. Данный вариант архитектуры сочетает в себе принцип централизованной сети и ad-нос сети. Она подходит для сценария с большим количеством БПЛА и различными характеристиками полета или характеристиками каналов связи. Поскольку в ней частично используется централизованный принцип, то эта архитектура связи в ряде случаев, не обеспечивает достаточной связности. В рамках отдельных групп вероятность связности БПЛА с центром управления, а также связности между БПЛА в группе аналогичны соответствующими показателями для ad нос сети, т.е. определяются выражениями (3) и (4). Связь между БПЛА различных групп возможна только через центр управления (рис.4).

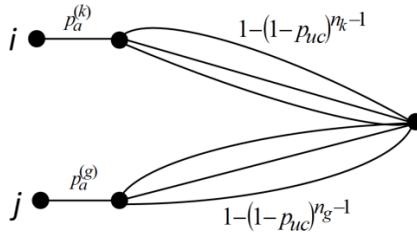


Рис. 4. Структура маршрута БПЛА – БПЛА в многогрупповой сети

Вероятность связности между БПЛА различных групп будет определяться как

$$p_a^{(kg)} = p_a^{(k)} p_a^{(g)} (1 - (1 - p_{uc})^{n_k - 1} (1 - p_{uc})^{n_g - 1})$$

где $p_a^{(k)}$ и $p_a^{(g)}$ вероятности связности в группах k и g ;
 n_k и n_g – количество БПЛА в группах k и g ;

p_{uc} – вероятность связности БПЛА с центром управления (наземной станцией).

Многослойная БПЛА ad-нос-сеть. Структура многослойной сети похожа на структуру многогрупповой ad-нос сети, за исключением того что имеют место непосредственные связи между разными группами БПЛА, без участия центра управления. Сеть верхнего слоя БПЛА представляет собой ad-нос сеть, которая состоит из головных

БПЛА всех групп. Для обмена данными между любыми двумя группами БПЛА не требуется пропускать данные через наземную станцию. Наземная станция получает данные, которые предназначены только для нее, что приводит к значительному снижению требований к ее пропускной способности. Многослойная БПЛА ad-hoc-сеть является наиболее надежной, поскольку не содержит единой точки, приводящей к отказу всей сети. В рамках отдельных групп вероятность связности БПЛА с центром управления, а также связности между БПЛА в группе аналогичны соответствующими показателям для ad hoc сети, т.е. определяются выражениями (3) и (4). Связь между БПЛА различных групп возможна как через центр управления, так и напрямую (рис.5).

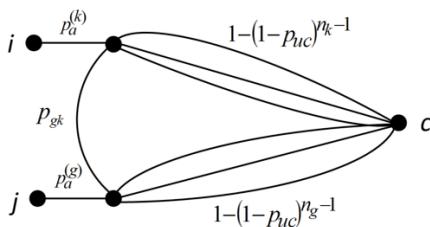


Рис. 5. Структура маршрута БПЛА – БПЛА в многослойной сети

Вероятность связности между БПЛА различных групп будет определяться как

$$p_u^{(kg)} = p_a^{(k)} p_a^{(g)} (1 - (1 - p_{kg})(1 - (1 - p_{uc})^{n_k - 1}(1 - p_{uc})^{n_g - 1}))$$

где $p_a^{(k)}$ и $p_a^{(g)}$ вероятности связности в группах k и g ;

n_k и n_g – количество БПЛА в группах k и g ;

p_{uc} – вероятность связности БПЛА с центром управления;

p_{kg} – вероятность связности между группами k и g .

Сравнительная характеристика. По сравнению с централизованной структурой, децентрализованные сети являются более гибкими, надежными и имеют лучшие эксплуатационные характеристики. В децентрализованной сети БПЛА связаны между собой непосредственно или через несколько транзитных участков (скачков). В таблице 1 приведено сравнение четырех основных архитектур построения сети связи для группы беспилотных летательных аппаратов. Полносвязные беспроводные сети имеют преимущества в части безопасности и надежности. В децентрализованной архитектуре особые требования предъявляются к протоколу маршрутизации, который должен обеспечить доставку пакетов от источника к узлу. Он должен обеспечивать

Таблица 1

Сравнение четырех базовой архитектуры построения сети связи для группы БПЛА

Характеристика	Централизованная	ad hoc сеть	Много-групповая ad hoc сеть	Много-слойная ad hoc сеть
Устойчивость	Низкая	Средняя	Средняя	Высокая
Связность	через ЦУ	маршрут в ad hoc сети	маршрут в ad hoc сети	маршрут в ad hoc сети
Размер сети	малая	средняя	большая	большая
Маршрут	один скачок	много скачков	много скачков	много скачков
Покрытие	ограничено связью с ЦУ	опр. числом БПЛА	опр. числом БПЛА	опр. числом БПЛА
Характер задач	Однор.	Однор.	Разнор.	Разнор.

выбор маршрута, отвечающего поставленным требованиям. Протокол маршрутизации обеспечивает самоорганизацию ad-hoc сети. Наземная станция также может обрабатывать данные для для поддержки решений по конфигурации сети [5].

3. Выводы

Анализ структур построения сети связи БПЛА позволи сделать следующие выводы:

1. Для построения сетей БПЛА могут быть использованы централизованная, децентрализованная и комбинированная архитектуры.

2. Применение централизованной архитектуры целесообразно при относительно малом количестве БПЛА и высокой вероятности связности БПЛА – центр управления.

3. Применение децентрализованной архитектуры, обеспечивает большую гибкость, надежность и лучшие эксплуатационные характеристики. Ее применение целесообразно при относительно большом количестве БПЛА и дает возможность повышения надежности связи,

как с центром управления, так и БПЛА между собой, а также расширения области действия группы БПЛА.

4. Многогрупповая архитектура построения сети обеспечивает связь между относительно большим числом БПЛА, при этом в рамках одной группы ее свойства аналогичны сети с децентрализованной архитектурой. Ее применение целесообразно при относительно большом числе БПЛА и малом трафике между группами.

5. Многослойная архитектура является наиболее универсальной, она обеспечивает связность, как в группах, так и между группами БПЛА. Ее применение целесообразно при относительно большом числе БПЛА и необходимости организации связи между группами.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта N15-07- 09431а "Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей".

Литература

1. *Кучерявый А. Е.* Летающие сенсорные сети / А. Е. Кучерявый, А. Е. Владыко, Р. В. Киричек, А. И. Парамонов, А. В. Прокопьев, А. И. Богданов, А. А. Дорг-Еольц // Электросвязь. - 2014. - № 9.
2. *Дьяченко.* Задача формирования строя в группе БПЛА / Александр Александрович // Известия Южного федерального университета. Технические науки 128.3 - 2012.
3. *Li, Jun.* Communication architectures and protocols for networking unmanned aerial vehicles / Yifeng Zhou, Lisa Lamont // Globecom Workshops (GC Wkshps), 2013 IEEE. IEEE, 2013. p. 1415-1420.
4. *Gupta.* Survey of Important Issues in UAV Communication Networks / Lav, Raj Jain, and Gabor Vaszkun.
5. *I. Dalmaso.* WiMAX networks for emergency management based on UAVs / I. Galletti, R. Giuliano, F. Mazzenga // IEEE First AESS European Conference on Satellite Telecommunications (ESTEL), 2012, pp. 1-6

UDC 621.395

Analysis of communications network based unmanned aerial vehicles

T. N. Dao, A. I. Paramonov

Federal State Budget-Financed Educational Institution of Higher Education The Bonch-Bruевич Saint - Petersburg State University of Telecommunications

22-1, Prospekt Bolshevikov, 193232, Saint-Petersburg, RUSSIA

The article presents the results of the analysis methods for constructing networks based on unmanned aerial vehicles (UAVs). The main options for the structure of communication networks formed by a group of UAVs as between the UAV in the group, still between the UAV and the control center. The main features and characteristics of the various ways of networking, their advantages and disadvantages. The results obtained make it possible to choose the best way to organize the network of UAV based on the requirements defined by the network purpose and its parameters such as the number of UAVs, the number of groups, the network scale, the characteristics of the movement.

Keywords: unmanned aerial vehicles, centralized network, decentralized network.

УДК 004.4

Моделе-ориентированная проверка правильности сетевого взаимодействия

В. В. Десятков, М. Т. Наунг

*Кафедра информационных систем и телекоммуникаций,
Московского государственного университета им. Н.Э.Баумана,
ул. 2-ая Бауманская, д. 5, Москва, Россия, 105005*

Аннотация. Настоящая статья посвящена развитию методики формальной проверки правильности сетевого взаимодействия. В качестве языка для решения задач формальной проверки правильности сетевого взаимодействия выбрана логика тайлов, а в качестве языка для формулировки этих свойств - временная модальная логика. В статье обосновывается такой выбор, рассматриваются принципы проверки правильности (доказательства) сетевого взаимодействия как проверки правильности взаимодействия агентов в мультиагентной системе. Автоматизацию проверки правильности взаимодействия предлагается осуществлять логическими программами, получаемыми в результате перехода от описания сетевого взаимодействия в логике тайлов и требований правильности взаимодействия на языке модальной логики к программе проверки правильности на языке логического программирования Visual Prolog. Методика формальной проверки правильности сетевого взаимодействия иллюстрируется примером получения описания поведения агентов в логике тайлов, свойства обязательной реакции в модальной логике и логической программой для случая трех взаимодействующих агентов: руководителя и двух исполнителей. Обсуждаются перспективы использования предлагаемой методики для других свойств и приложений.

Ключевые слова: сетевое взаимодействие, логика тайлов, модальная логика, проверка правильности, интеллектуальные агенты, язык логического программирования ПРОЛОГ..

1. Введение

В работе [1] рассматривались свойства мультимодальных интеллектуальных интерфейсов, реализация которых, по мнению авторов, обеспечивает естественное «безбарьерное» взаимодействие пользователя с машиной. Какие из этих свойств необходимо воплощать, конечно, зависит от конкретной системы, где требуется соответствующий интерфейс. В любом случае, если какое-либо свойство интерфейса должно быть реализовано, то, во-первых, необходима его точная спецификация, а во-вторых, возможность формальной проверки этой спецификации применительно к интерфейсу конкретной системы. Поскольку речь идет, как правило, о программной системе, то, по существу, упомянутая спецификация свойств интерфейса и их формальная проверка, может рассматриваться как специфическая область проверки правильности сетевого взаимодействия программных систем в целом. К настоящему времени известно три основных подхода к проверке

правильности программных систем: тестирование наличия правильности, формальное доказательство правильности в рамках какого-либо исчисления и модели-ориентированное доказательство правильности. Различие между этими подходами состоит в следующем. При тестировании на вход модели подаются заранее подготовленные и представленные в каком-либо языке последовательности действий (восприятий и реакций агентов) и проверяется, проявляет ли модель правильное поведение. Известным недостатком тестирования является то, что мощность множества тестирующих последовательностей действий (нитей) может быть слишком велика, что приводит к комбинаторному взрыву и вычислительным проблемам тестирования. Кроме того, множество тестирующих последовательностей, как правило, формируется экспертом, и нет никакой гарантии полноты этого множества для проверки правильности. При формальном доказательстве наличия правильного поведения в рамках какого-либо исчисления эти свойства формулируются в виде теорем (целей) на языке соответствующего исчисления, которые затем доказываются (выводятся) из начальных знаний (аксиом), представленных на том же языке, используя специальные правила вывода. Если стратегия вывода полна, то успех доказательства наличия правильности зависит от полноты аксиоматизации начальных знаний. Кроме того, полная стратегия вывода может оказаться вычислительно сложной. Но главное, что ограничивает использование исчислений, – это сложность интерпретаций и аксиоматизации на языке этих исчислений необходимых сведений участниками процесса проверки правильности, которые, как правило, не являются специалистами в области исчислений. Модели-ориентированное доказательство наличия правильности, в отличие от формального доказательства в рамках исчисления, использует модели архитектуры и поведения системы, в которой реализуется взаимодействие. Если создатель этих моделей считает, что, создавая архитектуру и описывая поведение системы, он выполнил все предъявляемые к ним требования, то для проверки правильности модели остается только доказать, что ее архитектура и поведение этим требованиям удовлетворяют. Для подобного доказательства здесь используются правила вывода, но семантика этих правил существенно ориентирована на привычные для тех или иных приложений понятия и смыслы. Настоящая работа основана как раз на модели-ориентированном доказательстве свойств систем. Одними из самых популярных моделей, используемых для этих целей, являются процессные модели описания архитектуры и поведения систем. Прародителем целого класса таких моделей является, например, язык процессных выражений процессной алгебры (пи-исчисления) Роберта Милнера [2]. Известно, что процессные алгебры предлагают достаточно естественный путь описания параллельных систем, состоящих из агентов, взаимодействующих по общим каналам. Описание динамического поведения агентов обычно использует операционную семантику, представляемую в виде правил. К сожалению, языки процессных алгебр хотя и изящны с математической точки зрения, но

зачастую слишком абстрактны для применения, а управление правилами вывода в этих языках требует дополнительных усилий, не всегда следующих непосредственно из контекста. В связи с этим популярным является использование так называемых языков архитектурного описания (Architecture Description Language – ADL) [3], которые позволяют на разных уровнях абстракции описывать структуру и поведение систем. В настоящее время предложено довольно большое число различных языков архитектурного описания [4] [5] [6] [7], рассчитанных прежде всего на описание и проверку программных систем. В настоящей статье для описания поведения систем и доказательства правильности взаимодействия предлагается использовать формальные системы, основанные на правилах, называемых тайлами (tile). На русский язык термин tile переводится как черепица, плитка, изразец, что не очень естественно использовать для наших целей. Поэтому в настоящей статье будем использовать англицизм «логика тайлов» (tile logic). Принципы доказательства каких-либо свойств информационных систем на основе логик тайлов известны достаточно давно [8]. Но для того, чтобы какую-либо из этих логик можно было использовать для того или иного приложения, необходимо создать множество тайлов, соответствующих этому приложению. Нашим приложением является проверка (доказательство) правильности взаимодействия. Сетевых программных систем. Дальнейшая структура материала статьи следующая. В следующем разделе будут приведены основные понятия, связанные с тайлами. Затем на простом примере проверки правильности взаимодействия, называемого обязательной реакцией, будет введено необходимое для этого примера множество тайлов. После этого будет показано, как это множество используется для доказательства обязательной реакции, сформулированного на языке модальной логики. В заключении обсуждаются направления дальнейших работ в области проверки правильности взаимодействия

2. Принципы проверки правильности сетевого взаимодействия

Тайлы и операции над ними. Достоинства логики тайлов связаны со структурой ее правил вывода, каждое из которых может быть представлено в виде тайла (tile), структура которого в общем виде показана на рис. 1.

Здесь α – имя тайла. Тайл позволяет описывать поведение частично определённых компонент системы, содержащих переменные и называемых конфигурациями. Это поведение описывается в терминах возможных взаимодействий с внешней или внутренней средой. Поведение системы в целом выглядит как скоординированное взаимодействие отдельных тайлов. Смысл обозначений на рис. 1 тайла следующий: $?i_i$ – начальный входной интерфейс тайла; $?s_i$ – начальная конфигурация тайла; $?o_i$ – начальный выходной интерфейс тайла; $!i_o$ – конечный входной интерфейс тайла; $!s_o$ – конечная конфигурация тайла, $!o_o$ –

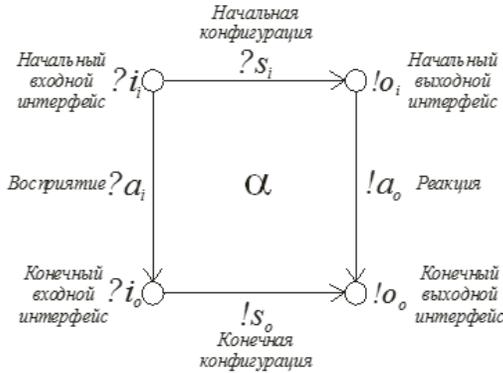


Рис. 1. Тайл

конечный выходной интерфейс тайла; $?a_i$ – восприятие тайла; $!a_o$ – реакция тайла. Восприятия и реакции называются также действиями. Интерфейсы тайла задают форматы его взаимодействия с внешним окружением. Тайл α на рис.1 в ответ на восприятие $?a_i$ задает преобразование начальной конфигурации $?s_i$ в конечную конфигурацию $!s_o$, производя реакцию $!a_o$. Такое преобразование возможно только тогда, когда тайл не только осуществил восприятие $?a_i$, но и когда форматы данных внешнего окружения тайла удовлетворяют форматам интерфейсов тайла.

Тайлы, у которых все интерфейсы одинаковы называются секвентами. Обозначения интерфейсов в секвентах опускаются. Графическое изображение секвента показано на рис. 2. и параллельно для того, чтобы задавать более сложное поведение систем. Операция горизонтальной композиции тайлов обозначается символом $*$. Горизонтальная композиция двух тайлов α и β показана на рис. 3.

Горизонтальная композиция двух тайлов α и β возможна только при условии, что $!a_{o\alpha} = !a_{i\beta}$, $!o_{i\alpha} = ?i_{i\beta}$, $!o_{o\alpha} = ?i_{i\beta}$. Смысл горизонтальной композиции тайлов заключается в моделировании формирования более сложных конфигураций и из более простых при условии выполнения равенств $!a_{o\alpha} = !a_{i\beta}$, $!o_{i\alpha} = ?i_{i\beta}$, $!o_{o\alpha} = ?i_{i\beta}$. Операция вертикальной композиции обозначается символом \bullet . Вертикальная композиция двух тайлов α и β показана на рис. 4.

Вертикальная композиция моделирует последовательное преобразование конфигурации $?s_{i\alpha}$ в $!s_{o\beta}$ соответственно сначала с помощью

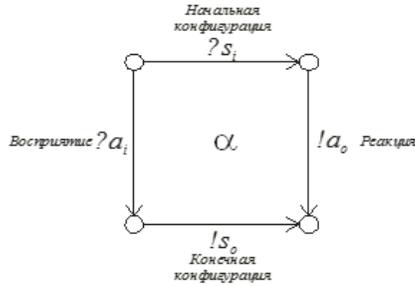


Рис. 2. Секвент

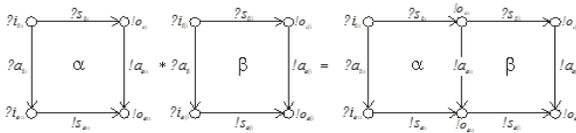


Рис. 3. Горизонтальная композиция тайлов

пары восприятие $?a_{i\alpha}$, реакция $!a_{o\alpha}$, а затем пары восприятие $?a_{i\beta}$, реакция $!a_{o\beta}$, и при условии, что $!s_{o\alpha} = ?s_{i\beta}$, $!o_{o\alpha} = ?i_{i\beta}$, $!o_{o\alpha} = ?i_{i\beta}$.

Операция параллельной композиции обозначается символом ?. Параллельная композиция тайлов показана на рис. 5.

Параллельная композиция двух тайлов α и β возможна только при условии параллельного выполнения соответственно пар восприятий $?a_{o\alpha}, ?a_{i\beta}$, реакций $!a_{o\alpha}, !a_{i\beta}$, входных конфигураций $?s_{o\alpha}, ?s_{i\beta}$, выходных конфигураций $!s_{o\alpha}, !s_{o\beta}$.

Та или иная логика тайлов получается в результате введения множества базовых тайлов и некоторой совокупности вспомогательных

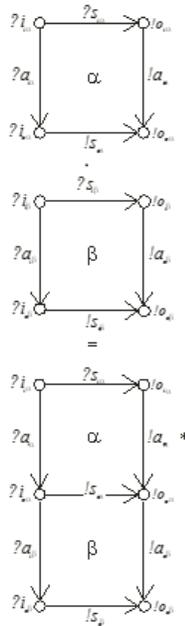


Рис. 4. Вертикальная композиция тайлов

тайлов, над которыми могут совершаться введенные выше операции, приводящие к различного рода преобразованиям тайлов, в частности, с целью доказательства свойств интеллектуальных интерфейсов.

2.1. Представление сетевого взаимодействия

Для проверки правильности сетевого взаимодействия (интерфейсов) больших и сложных систем требуется представлять систему как иерархическое множество независимых процессов (агентов) и каналов связи между этими процессами, которые могут реконфигурироваться. Логика тайлов в наибольшей степени подходит для проверки правильности сетевого взаимодействия больших и сложных систем, поскольку интерфейсы являются базовыми объектами, которыми эта логика способна манипулировать. Различные схемы соединений интерфейсов могут создаваться в логике тайлов благодаря параллельной

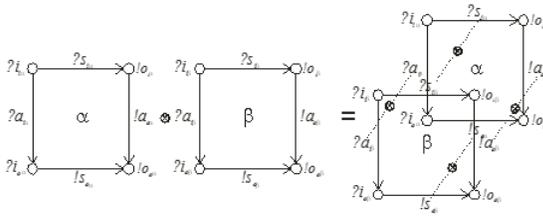


Рис. 5. Параллельная композиция тайлов

и горизонтальной композициям, порождающим различные системные конфигурации, каждая из которых задает пространственное представление общего состояния системы, состоящего из локальных состояний. Горизонтальная или параллельная композиции обеспечивают также возможность контекстной зависимости данных благодаря интерфейсам. Поведение системы в целом определяется множеством тайлов, задающих локальные изменения конфигураций системы, являющихся локальными состояниями системы. Для того, чтобы сказанное было более понятным, введем следующие обозначения. Архитектуру системы любого уровня иерархии будем представлять четверкой $S=P, L, I, O$, где

- P – множество процессов системы, изображаемых прямоугольниками,
- L – множество каналов системы, соединяющих процессы,
- I – множество подмножеств $I(p) \subset L$ входных каналов процессов $p \in P$,
- O – множество подмножеств $O(p) \subset L$ выходных каналов процесса $p \in P$.

Каждый процесс $p \in P$ верхнего уровня системы может задаваться системой более низкого уровня $S(p)=P(p), L(p), I(p), O(p), ei(p), eo(p)$, где $\bullet P(p), L(p), I(p), O(p)$ имеют тот же смысл, что и P, L, i, o , но только как внутренняя структура процесса p .

- $ei(p)$ – функция, отображающая внутренние каналы процесса p в его внешние входные каналы,
- $eo(p)$ – функция, отображающая внутренние каналы процесса p в его внешние выходные каналы.

Рассмотрим пример, иллюстрирующий предлагаемые принципы проверки правильности взаимодействия. Этот пример в некоторой степени использует пример клиент-серверной системы, заимствованной

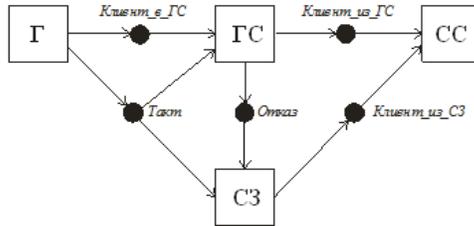


Рис. 6. Клиент-серверная система S

из работы [10]. Система состоит из главного сервера (ГС), отвечающего за интерфейс с клиентами, сервера задержки (СЗ), перераспределяющего клиентов в случае их длительного обслуживания главным сервером, генератора (Г), фиксирующего клиентов и направляющего их главному серверу, и сервера статистики (СС), фиксирующего обслуженных клиентов и сервера, их обслуживавшие. Клиенты, слишком долго ожидающие обслуживания главным сервером, могут отказаться от его услуги и обратиться к серверу задержек. Сервер задержек является конкурентом главного сервера и может оказывать те же услуги, что и главный сервер, но за более длительное время. Каждый клиент сам решает, следует ли ему отказываться от услуг главного сервера без их предоставления и переходить на обслуживание к серверу задержек. или нет. Пример архитектуры клиент-серверной системы S верхнего уровня графически представлен на рис. 6.

Согласно приведенным выше определениям для клиент-серверной системы первого уровня, представленной на рис. 6, будем иметь

- $P = \Gamma, ГС, СС, СЗ,$
- $L = \text{Клиент в ГС, Клиент из ГС, Такт, Отказ, Клиент из СС},$
- $I = I(ГС), I(СС), I(СЗ),$
 - $I(\Gamma) = \emptyset,$
 - $I(ГС) = \text{Клиент в ГС, Такт},$
 - $I(СС) = \text{Клиент из ГС, Клиент из СС},$
 - $I(СЗ) = \text{Такт, Отказ},$
- $O = O(\Gamma), O(ГС), O(СЗ),$
 - $O(\Gamma) = \text{Клиент в ГС, Такт},$
 - $O(ГС) = \text{Клиент из ГС, Отказ},$

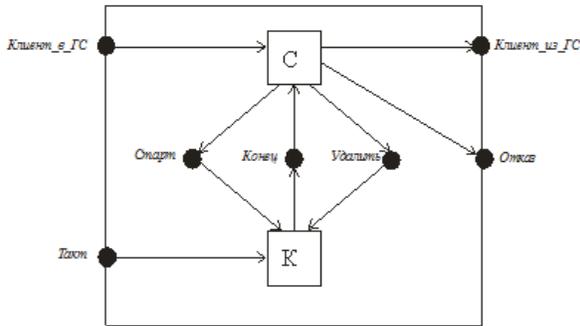


Рис. 7. Архитектура главного сервера (ГС)

- o $O(CC) = \emptyset$,
- o $O(CЗ) = \text{Клиент из } CЗ$.

Пример архитектуры второго уровня для главного сервера на рис. 6 показан на рис. 7. Здесь С - буфер, К - клиентский сервис.

Для архитектуры главного сервера, представленной на рис. 7 будем иметь

- $P(ГС) = C, K$,
- $L(ГС) = \text{Старт, Конец, Удалить}$,
- $I(ГС) = I(C), I(K)$,
- o $I(C) = \text{Клиент в } ГС, \text{ Конец}$
- o $I(K) = \text{Такт, Старт, Удалить}$
- $O(ГС) = O(C), O(K)$,
- o $O(C) = \text{Клиент из } ГС, \text{ Отказ}$
- o $O(K) = \text{Конец}$
- $ei(ГС) = I(ГС) = \text{Клиент в } ГС, \text{ Такт}$
- $eo(ГС) = O(ГС) = \text{Клиент в } ГС, \text{ Отказ}$

Для простоты тайлы будем именовать именами процессов (агентов), конфигурации тайлов будем обозначать подмножествами их каналов, а действия над ними по помещению и получению конкретных сообщений в каналы, - подмножествами имен каналов с указанием действий поместить (put) или получить (receive), совершаемых над ними.

язык временной модальной логики. Поскольку этот язык хорошо изучен и его применение для формулировки правильности также широко известно, то мы в настоящей статье не будем уделять ему слишком много внимания. Рассмотрим лишь принципы формулировки условий правильности взаимодействия с привязкой к понятиям тайлов, с помощью операций над которыми эти условия будут проверяться. В качестве примера одного из таких свойств выберем простейшее свойство обязательной реакции. В простейшем случае суть этого свойства применительно к взаимодействию, например, агента-руководителя АР и какого-либо агента-исполнителя АИ состоит в следующем. На каждую реакцию агента-руководителя АР, в результате которой в его выходном канале x , являющемся одновременно входным каналом для агента-исполнителя АИ, окажется какое-либо сообщение, предназначенное для последнего, этот агент-исполнитель АИ обязательно должен получить эти сообщение и реагировать на него помещением в какой-либо выходной канал u , являющийся одновременно входным для агента-руководителя, ответного сообщения. На языке модальной логики формулировка этого условия для нашего простейшего примера может выглядеть, например, следующим образом:

$$\begin{aligned} & \square[[tile(\alpha_{AP}^1, ?s_{AP}^1, f(put(AP(x))), !s_{AP}^1)] \\ & \supset [tile(\alpha_{AU}^1, ?s_{AU}^1, f(receive(AU(x))), !s_{AU}^1)] \\ & \wedge [tile(\alpha_{AU}^2, ?s_{AU}^2, f(put(AU(x))), !s_{AU}^2)] \\ & \supset [tile(\alpha_{AP}^2, ?s_{AP}^2, f(receive(AP(x))), !s_{AP}^2)]]]. \end{aligned}$$

Здесь каждый предикат $tile(\alpha, ?s_i, f(), !s_o)$ представляет тайл, где α – имя тайла, $f()$ – функтор, структура которого может быть достаточно сложной в зависимости от структуры восприятий и реакций, $?s_i, !s_o$ – конфигурации.

Очевидно, что для реальных систем большой размерности, включающих n взаимодействующих агентов, каждый из которых связан друг с другом m каналами, генерация наборов восприятий и реакций и проверка на них требуемых условий правильности вручную трудно осуществима и требуется автоматизация подобной проверки. Перспективным путем автоматизации является использование логического программирования, например, на основе языка логического программирования VISUAL PROLOG [9]. Структура подобной логической программы проверки правильности взаимодействия в простейшем случае включает, как минимум разделы domains, predicates, clauses, goal. Раздел domains, который содержит описание структуры всех используемых функторов. Раздел predicates, который содержит описание предикатов, представляющих все тайлы, участвующие в проверке правильности взаимодействия и начинающиеся с префикса tile.

Раздел *clauses*, который содержит необходимые факты и правила, полученные в соответствии с условиями правильного взаимодействия, сформулированными на языке модальной логики. Раздел *goal*, который содержит проверяемое условие правильности взаимодействия.

3. Заключение

В настоящей статье изложены принципы модели-ориентированной проверки правильности сетевого взаимодействия на основе использования тайлов. Вследствие краткости статьи представить в достаточно представительном объёме все возможности, которые открывает использование тайлов для анализа правильности взаимодействия затруднительно. Следует заметить, что правильный выбор и полнота множества базовых тайлов позволяют 1) эффективно управлять доказательством правильности взаимодействия, требования к которому формулируются на языке модальной логики; 2) тестировать правильность взаимодействия в процессе перехода от одних конфигураций к другим при выполнении множеств нитей с помощью вертикальной композиции тайлов; 3) тестировать правильность иерархического и контекстно зависимо взаимодействия в процессе перехода от одних конфигураций к другим при выполнении множеств нитей с помощью горизонтальной композиции тайлов; 4) тестировать правильность параллельного взаимодействия в процессе перехода от одних конфигураций к другим при выполнении множеств нитей с помощью параллельной композиции тайлов. Описанная в статье методика реализована в языке логического программирования *Visual Prolog* и показала свою высокую эффективность .

Литература

1. *Десятков В.В., Алфимцев А.Н.* Необходимые и достаточные формальные свойства мультимодального интерфейса // Вестник МГТУ им. Н.Э.Баумана. Сер. Приборостроение. Спец. вып. «Информационные технологии и компьютерные системы». М., 2011. С. 159-167.
2. *Miller R.* *Communicating and Mobile Systems: The π -calculus.* Cambridge.- UK: Cambridge University Press, 2003. 159 с.
3. *Medvidovic N., Taylor R. M.* A Classification and Comparison Framework for Software Architecture Description Languages // IEEE Transactions on Software Engineering, Vol 26, no 1, Janvier, 2000. pp 70-93.
4. *Braga C., Sztajnberg A.* Towards a Rewriting Semantics for a Software Architecture Description Language // Proceedings of WMF 2003, 6th Workshop on Formal Methods, Campina Grande, Brazil, E.N.T.C.S. 95, 2003. pp. 148-168.

5. *Bruni R., Fiadeiro J. L., Lanese I., Lopes A., Montanari.* New Insight into the Algebraic Properties of Architectural Connectors // IFIP TCS, 2004. pp. 367-380.
6. *Choutri A., Belala F., Barkaoui K.* A Tile Logic Based Approach for Software Architecture Description Analysis // USA., J. Software Engineering and Applications, 3, 2010. pp. 1067-1079.
7. *Bouanaka C., Choutri A., Belala F.* On Generating Tile System for a Software Architecture : Case of a Collaborative Application Session. // ICISOFT2007 (the Second Conference on Software and Data Technologies), July 22-25, 2007. pp. 123-128.
8. *Bruni R.* Tile Logic for Synchronized Rewriting of Concurrent Systems. Phd Thesis, University of Pisa, TD-1/99, March 1999.
9. *Адаменко А.Н., Кучков А.* Логическое программирование и Visual Prolog. Изд. БХВ– Петербург, 2003.
10. *Bouanaka C. Belala F. Barkaoui K.* A Tile Logic Based Semantics for Mobile Software Architectures. // International Journal of Critical Computer-Based Systems 01, 2(3), 2011. pp 288-308.

UDC 004.4

Model-oriented check the correctness of network interaction

V. V. Devyatkov, M. T. Naung

*Department of Information Systems and Telecommunications
Bauman Moscow State Technical University
2-nd Baumanskaya st. 5, Moscow, 105005, Russia*

In this article we developed the methodology for formal checking validity of interfaces. We selected tiles logic as a language for solving problems of checking validity of interfaces and temporal modal logic as the language for validity demands formulation. of checking validity (validity proof) of the interfaces as the validation of the agents' interaction in multi-agent system. The article explains the choice, considers the principles of validity proof of the interfaces as the validation of the agents interaction in multi-agent system. For automation of the verification we use the logic programs. The transition from the description of agents' interaction in tile logic and validity demands formulation in modal logic in logic programming language Visual Prolog are considered. The methodology is illustrated by a description of the agent's behavior in the tiles logic, validity demands formulation in the modal logic and logic program structure. The prospects of using the proposed methodology and applications is discussed.

Keywords: network interface, logic tiles, modal logic, validity demands formulation, intelligent agents, logic programming language Prolog.

УДК 004.73

Разработка и исследование методов инсталляции беспроводных сенсорных узлов с беспилотного летательного аппарата

Ч. З. Динь, Р. В. Киричек

*Кафедра сетей связи и передачи данных,
Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М.А.Бонч-Бруевича,
пр. Большевиков д.22, корп. 1, Санкт-Петербург, Россия, 193232*

Аннотация. В статье представлены разработка и сравнение способов инсталляции сенсорных узлов для покрытия заданной области при использовании различных способов: случайного размещения и детерминированного размещения. Рассмотрены методы повышения энергоэффективности сенсорных узлов и квадрокоптера. Исследованы методы инсталляции сенсорных узлов с использованием БПЛА различного типа.

Ключевые слова: всепроникающие сенсорные сети, летающая сенсорная сеть, беспилотный летательный аппарат, беспроводной сенсорный узел, инсталляция, покрытие.

1. Введение

В настоящее время развитие сетей связи осуществляется на основе концепции Интернета Вещей [1, 2]. Технологической базой для реализации концепции Интернета Вещей являются беспроводные сенсорные сети [3, 4]. Развитие технологий беспроводных сенсорных сетей привело к их повсеместному внедрению и появлению термина всепроникающие сенсорные сети (ВСС) (Ubiquitous Sensor Networks - USN). Такие сети стали одной из основных технологических платформ для Интернета Вещей.

В последние годы исследования в области ВСС постепенно переходят от изучения характеристик на плоскости к моделям в трехмерном пространстве. Одними из новых областей применения технологий ВСС являются летающие сенсорные сети (ЛСС). При этом в качестве летающих узлов используются общедоступные беспилотные летающие аппараты общего назначения (БПЛА) [5].

В статье проводится разработка и исследование методов инсталляции беспроводных сенсорных узлов с беспилотного летательного аппарата. Проводится обзор полного жизненного цикла построения летающей сенсорной сети и конкретизируются задачи, ориентированные на гарантированное покрытие заданной области сенсорами с беспилотного летательного аппарата. В работе разработана математическая модель и имитационная модель распределения сенсорных узлов с БПЛА

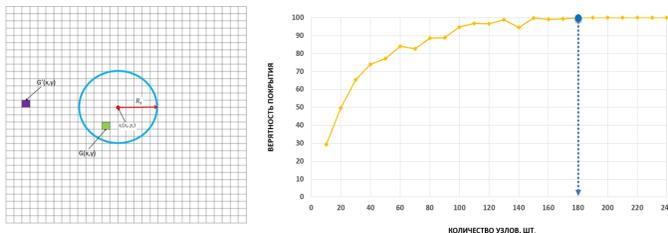


Рис. 1. Алгоритм и результат моделирования при случайной установке сенсорных узлов

самолетного и вертолетного типа. Полученные данные могут быть использованы для планирования летающих сенсорных сетей.

2. Основная часть

Летающая сенсорная сеть – строится на базе общедоступных беспилотных летающих аппаратов (БПЛА), которые применяются для сбора данных с сенсорных полей и доставки этих данных в сеть связи общего пользования [6, 7].

Летающая сенсорная сеть предполагается два сегмента: наземный и летающий, которые взаимодействуют между собой. Одна из первостепенных задач в летающих сенсорных сетях заключается в том, что БПЛА необходимо установить сенсорные узлы на местности (наземный сегмент) для покрытия территории и оптимального сбора информации. В нашей работе, используем квадрокоптер – один из видов БПЛА. Как известно, существует два способа установки узлов в беспроводной сенсорной сети: случайное размещение и детерминированное размещение [8]. Сенсорные узлы могут быть случайным образом разбросаны по некоторой территории или предварительно размещены в соответствии с планом построения сети. Естественно, что алгоритмы распределения данных между сенсорными узлами в первом и втором случае могут быть существенно различны.

В данной работе [9], рассмотрена вероятность покрытия при случайном размещении сенсорных узлов на заданной территории. Мы использовали двоичную модель [10], в которой модели, каждый сенсорный узел имеет одинаковый и постоянный радиус действия сенсора. Область обслуживания накроем сеткой линий $n \times m$, с размером каждой ячейки сетки равен 1. Для описания покрытия узлов сетки $G(x, y)$ введем двоичную функцию, значение которой равно 1, если узел попадает в зону покрытия сенсора $s_i(x_i, y_i)$ и 0, в

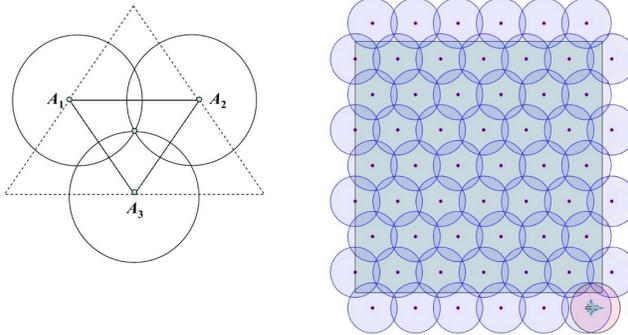


Рис. 2. Алгоритм и результат моделирования детерминированной инсталляции сенсорных узлов.

противном случае.

$$I(x, y, s_i) = \begin{cases} 1, & \sqrt{(x - x_i)^2 + (y - y_i)^2} < R_s \\ 0, & \end{cases}$$

Для анализа зависимости доли покрытия зоны обслуживания от числа сенсорных узлов была разработана имитационная модель. В области с размером 500 м на 500 м размещено $k = 0 \dots 240$ узлов с радиусом покрытия $R = 50$ м. Точки размещения узлов определяются координатами (x_i, y_i) , $i = 1 \dots k$ где x_i и y_i случайные числа, имеющие равномерное распределение в интервале от 0 до 100. Результат моделирования приведен на рис. 1.

Также был рассмотрен метод детерминированной инсталляции сенсорных узлов на заданной территории [11]. В данном методе используются правильные многоугольники, в которых сенсорные узлы размещаются в вершинах равносторонних треугольников [12–14]. Для оценки эффективности моделей покрытия, была разработана имитационная модель на базе пакета Anylogic. В области с размером 500 м на 500 м размещены сенсорные узлы с радиусом покрытия $R = 50$ м, для сбора данных используется квадрокоптер со средней скоростью 10 м/с. Результат моделирования приведен на рис. 2.

Из результатов видно, что при случайном размещении (180 узлов), для обеспечения близкого к 100% покрытия требуется более, чем втрое больше узлов, чем при их детерминированном размещении (58 узлов). Однако, стоит отметить, что при том же числе узлов (58) при случайном их распределении обеспечивается более 80% покрытия

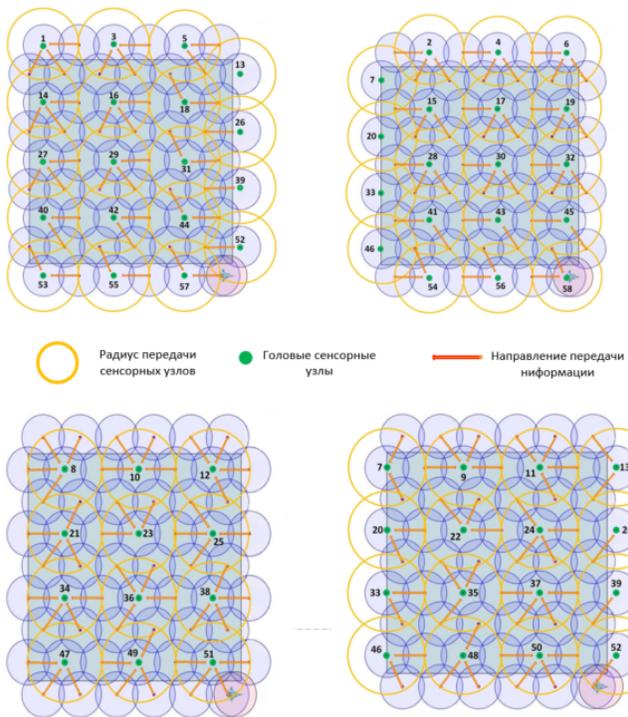


Рис. 3. Алгоритм выбора головных узлов для детерминированного размещения сенсорных узлов в сенсорной области с использованием квадрокоптера

территории. При детерминированном размещении узлов 100% покрытие достигается лишь в идеальном случае, на практике имеет место ошибка их позиционирования, которая приводит к уменьшению доли покрытия. С учетом последнего можно полагать, что оба метода сопоставимы по эффективности и должны выбираться исходя из конкретных условий.

Наряду с задачей размещения сенсорных узлов для покрытия максимального объема территории, актуальной задачей построения БСС является задача повышения эффективности энергопотребления. При использовании квадрокоптеров, необходимо обеспечить экономию не только энергии наземного сегмента сети, но и также энергию квадрокоптера.

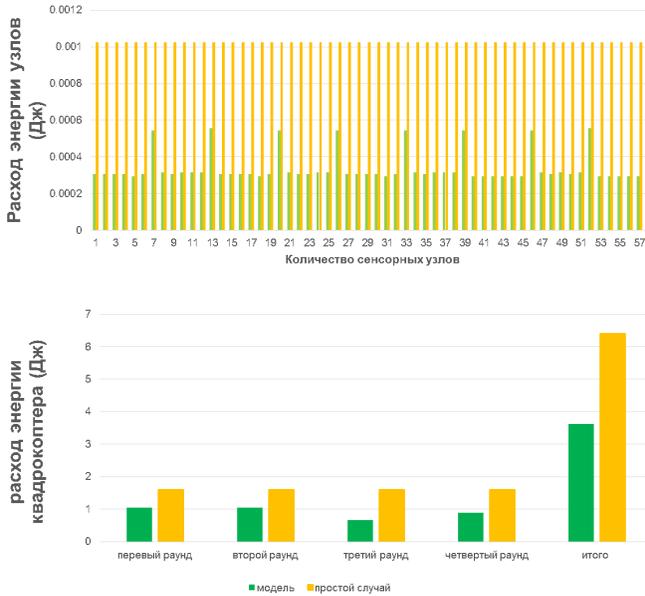


Рис. 4. Результаты моделирования расхода энергии сенсорных узлов и квадрокоптера

На основе работ [9, 11] был разработан алгоритм выбора головных узлов кластеров при детерминированном размещении сенсорных узлов в области обслуживания и сбора информации с них с помощью квадрокоптера. Данный алгоритм предполагает, что каждый из узлов сети может выполнять функции головного узла. Эти функции делегируются ему периодически, так чтобы каждый из узлов сети периодически становился головным узлом. Работа алгоритма проиллюстрирована на рис. 3. Головной узел кластера выделен зеленым цветом, члены кластера связаны с ним оранжевыми линиями. На этом рисунке приведены 4 раунда работы сети (смены роли узлов). За 4 раунда каждый из узлов сети единожды выполняет роль головного узла.

На рис. 4 приведены результаты моделирования расхода энергии при опросе каждого из узлов сети и при использовании алгоритма кластеризации и опросе только головных узлов кластеров. Как видно из представленных результатов, предлагаемая модель обеспечивает снижение затрачиваемой на каждом цикле движения квадрокоптера энергии и энергии сенсорных узлов почти вдвое.

Таблица 1
Сравнение БПЛА различного типа для решения задачи размещения узлов сенсорной сети

	БПЛА самолетного типа	БПЛА вертолетного типа
Продолжительность полета, мин.	60	30
Скорость полета, км/ч	40 - 100	20 - 60
Минимальная область покрытия	555м x 555м	275м x 275м
Максимальная область покрытия	880м x 880м	480м x 480м

Далее в данной работе было рассмотрено применение БПЛА различного типа для инсталляции узлов наземного сегмента, с помощью:

1. БПЛА самолетного типа. Имеет приблизительную точность инсталляции сенсорных узлов, вследствие постоянного движения БПЛА данного типа;

2. БПЛА вертолетного типа (мультикоптер). Имеет высокую точность инсталляции, в связи с характером движения БПЛА данного типа.

Сравнение данных БПЛА для решения задачи размещения узлов сенсорной сети приведены в таблице 1. По данной таблице видно, что БПЛА вертолетного типа больше подходят для более точного размещения сенсорных узлов на небольшой площади, тогда как БПЛА самолетного типа подходят для случайного размещения сенсорных узлов на большой площади.

3. Заключение

В результате данной работы было определено, что детерминированный способ размещения узлов, при условии достаточно точного их размещения обеспечивает большую эффективность при ограниченной точности размещения узлов, а также при менее жестких требованиях к доле покрытия. Было установлено, что случайное размещение узлов может быть сопоставимо по эффективности с детерминированным. Был разработан метод организации беспроводной сенсорной сети при детерминированном размещении узлов и использовании квадрокоптера для выполнения операций сбора данных. Предложенный метод

кластерной организации сети позволяет существенно снизить энергозатраты квадрокоптера на работу по сбору данных. Также были проанализированы характеристики каждого из типов БПЛА и получены результаты, описывающие случаи, при которых оптимально использовать БПЛА самолетного или вертолетного типа.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта N15-07-09431a "Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей".

Литература

1. *А.Е.Кучерявый* Интернет Вещей // Электросвязь. - 2013. - № 1. с.21-24.
2. *Росляков А. В., и др.* Интернет Вещей. — Самара: ШУТИ, 2014.
3. *Кучерявый А. Е., Парамонов А. И., Кучерявый Е. А.* Сети связи общего пользования. Тенденции развития и методы расчёта. — ФГУП ЦНИИС, 2008.
4. *Кучерявый А. Е., Кучерявый Е. А.* От e-России к u-России: тенденции развития электросвязи. — Электросвязь, 2005, № 5.
5. *Кучерявый А. Е., Владыко А. Г., Киричек Р. В., Парамонов А. И., Прокопьев А. В., Богданов И. А., Дорт-Гольц А. А.* Летающие сенсорные сети // Электросвязь. — 2014. — № 9. — С. 2-5.
6. *Киричек Р. В., Парамонов А. И., Прокопьев А. В., Кучерявый А. Е.* Эволюция исследований в области беспроводных сенсорных сетей // Информационные технологии и телекоммуникации. — 2014. — № 4 (8). — С. 29-41.
7. *Киричек Р. В., Владыко А. Г., Захаров М. В., Кучерявый А.Е.* Модельные сети для Интернета Вещей и программируемых сетей // Информационные технологии и телекоммуникации. — 2015. — № 3(11). — С. 17-26.
8. *Динь Ч. З., Киричек Р. В., Кучерявый А. Е.* Обзор методов инсталляции сенсорных узлов с квадрокоптера. — Информационные технологии и телекоммуникации, 2015, №1, С.50-61.
9. *Динь Ч. З., Киричек Р. В., Парамонов А. И., Кучерявый А. Е.* Имитационная модель инсталляции сенсоров с квадрокоптера на заданной территории // Информационные технологии и телекоммуникации, 2015, №2, С.93-100.
10. *Mohammed A. Z., Sabah M. A., Nabil S.* Coverage maximization in mobile Wireless Sensor Networks utilizing immune node deployment algorithm. — Electrical and Computer Engineering (CSECE), 2014 IEEE 27th Canadian Conference on.
11. *Динь Ч. З., Киричек Р. В., Парамонов А. И., Кучерявый А. Е.* Методы инсталляции сенсорных узлов с квадрокоптера

- общего пользования. — Информационные технологии и телекоммуникации, 2015, №4, С.57-67.
12. *Tom L. Ф.* Расположения на плоскости, на сфере и в пространстве. — Государственное издательство физико-математической литературы — 1958.
 13. *Wu J., Yang S.* Energy-efficient node scheduling models in sensor networks with adjustable ranges. — Int. J. Found. Comput. Sci. 2005, 16, 3-17.
 14. *Zalyubovskiy V., Erzin A., Astrakov S., Choo H.* Energy-efficient Area Coverage by Sensors with Adjustable Ranges. — Sensors 2009, 9, 2446-2460.

UDC 004.73

Development and research of methods of installation of wireless sensor nodes from UAV

T. D. Dinh, R. V. Kirichek

*Department of communications networks and data transmission
The Bonch-Bruевич Saint-Petersburg state university of communication
22 - 1, Prospekt Bolshevikov, 193232, Saint-Petersburg, RUSSIA*

In this article, we present the development and comparison of methods of installation of sensor nodes to cover a given area using different methods: random and deterministic placements. We study methods for improving the energy efficiency of sensor nodes and quadcopter. The methods of installation of sensor nodes using UAVs of different types have been researched.

Keywords: Ubiquitous Sensor Networks, flying sensor networks, UAVs, wireless sensor node, installation, coverage.

УДК 004.7

Обзор возможных видов и методов тестирования Интернет Вещей

Р. А. Долгушев, Р. В. Киричѐк

*Кафедра сетей связи и передачи данных,
Санкт-Петербургский Государственный Университет Телекоммуникаций
им. проф. М. А. Бонч-Бруевича,
пр. Большевиков, д.22, Санкт-Петербург, Россия, 193232*

Аннотация. Статья посвящена обзору видов и методов тестирования Интернет Вещей. Проводится анализ существующих методов тестирования, которые используются в сетях связи в настоящее время. На основе обзора предложен комплекс тестовых спецификаций для проведения комплексного тестирования устройств Интернета Вещей. В результате тестирования возможно оценить различные критерии: вероятностно-временные параметры, связанные с обменом данными между Вещами, а также с использованием сетевых ресурсов, протоколов и т.д. В качестве апробации предлагаемых тестовых спецификаций проведен натурный эксперимент по тестированию фрагмента сети связи с целью выбора прикладного протокола для типовой Интернет Вещи. Выбор протокола осуществляется на базе определения среднего времени отклика на запрос (RTT) при использовании протоколов MQTT, CoAP и HTTP.

Ключевые слова: Интернет Вещь, ИВ, IoT, ЛСС, FUSN, Ethernet, Wi-Fi, Bluetooth, ZigBee, MQTT, CoAP, HTTP.

1. Введение

В январе 2016 года на Всемирном экономическом форуме был отмечен тот факт, что в данный момент времени человечество находится на краю новой четвертой технологической революции (Industry 4.0), которая характеризуется слиянием технологий, стирающих грани между физическим и цифровым миром, а также появлением Интернета Вещей [1] и услуг, которые он может предложить.

Традиционно на протяжении долгих лет многие устройства тестировались изолированно, и никто не заботился о таких вещах, как пользовательский интерфейс (если таковой вообще был) или о подключении этих устройств к сети Интернет. Однако, в последние годы Интернет Вещей приобрел большую популярность, так как количество устройств, подключенных к сети Интернет неуклонно растет. Классический подход к тестированию сетевых устройств и активного сетевого оборудования не всегда может быть использован для тестирования Интернет Вещей. Нужно выработать новые стратегии, которые помогут добиться стабильной работы всех компонентов сети. Вследствие этого, вопрос об их тестировании является на данный момент открытым и

актуальным, поэтому появилась необходимость в разработке эффективных методов тестирования, которые помогли бы решить данную проблему.

Предполагается, что Интернет Вещи будут разделять вопросы разработки и тестирования во встраиваемых программных системах, также как это происходит и в традиционных IT- и web-системах. С ростом числа устройств Интернета Вещей и программных проектов потребность в тестировщиках и особых подходах к тестированию также возрастет. Специалисты в этой области столкнутся с множеством различных проблем и возможных ошибок.

Чтобы гарантировать качество продуктов и услуг Интернета Вещей необходимо выработать такие решения для тестирования, как:

1) Продуманная стратегия тестирования. Нужно сосредоточиться на эффективных методах и практиках тестирования. Четкие требования, детальный план испытаний должны быть частью данной стратегии, что позволит в дальнейшем выпустить качественный продукт или услугу;

2) Новые инструменты и платформы тестирования. Новые платформы будут необходимы для извлечения полезной информации из огромного количества необработанных данных, обеспечивая при этом основу для систем технической поддержки. Также понадобятся дополнительные инструменты, утилиты и симуляторы для тестирования, так как в противном случае данный процесс будет затруднен [2,3];

3) Тестирование методом «черного ящика». Для разработки эффективных тестов нужно уметь выявить архитектуру устройства, тип операционной системы, протоколы взаимодействия без знания внутренних механизмов тестируемого объекта.

2. Анализ методов и видов тестирования Интернет Вещей

Рекомендация МСЭ-Т Q.3900 «Архитектура модельной сети и методы тестирования технических средств сетей связи нового поколения (NGN) для использования в сетях связи общего пользования» [4], выпущенная в 2006 году, стала основным стандартом для тестирования NGN-сетей [5]. Данную концепцию можно использовать и относительно тестирования Интернет Вещей, однако, рекомендация Q.3900 не определяет структуру IoT-тестирования.

Как уже было сказано, тестирование Интернет Вещей так или иначе относится к проверке различных аспектов, касающихся функционирования самих устройств, их взаимодействия с сетью и безопасности [6]. Проблемы тестирования выходят за рамки устройств и датчиков, так как включают в себя дополнительные сложности, связанные с концепцией больших данных, т.е. большой объем и разнообразие генерируемой информации. Кроме того, на текущий момент, не существует единого подхода к тестированию устройств Интернета Вещей, которые существуют для узлов сети связи следующего поколения [2]. Таким



Рис. 1. Области тестирования Интернет Вещей

образом, создание модельной сети и различных методов тестирования поможет решить данную проблему.

Процесс тестирования крайне важен. Вследствие этого необходимо использовать специальные симуляторы, которые могут имитировать работу сетевых узлов и устройств Интернета Вещей, что позволит сократить расходы на оборудование. В целом, вопросы, касающиеся тестирования охватывают большую область, начиная с самих устройств, всевозможных датчиков [1] и заканчивая различными платформами, которые собирают и анализируют данные Интернет Вещей. Существует четыре основные области тестирования IoT-устройств, которые должны учитываться при разработке любого продукта, имеющего возможность подключения к сети Интернет [7].

Так как летающие сенсорные сети представляют собой одно из приложений концепции Интернета Вещей, то это позволяет применить области тестирования IoT-устройств и для ЛСС [8]. При построении летающих сенсорных сетей особое внимание необходимо обратить на безопасность, для того, чтобы не допустить вторжения в сеть, а также проверить подлинность узлов летающего и наземного сегмента. Приложения Интернета Вещей базируются на очень простых и дешевых узлах. Это может привести к клонированию элементов, поэтому необходимо изучить проблемы безопасности сетей FUSN.

Более подробно области тестирования приведены на рисунке 1.

Помимо тестирования устройств физического мира, также есть определенные виды тестирования виртуальных вещей. Виртуальные вещи существуют в информационном мире, их можно хранить, обрабатывать и получать к ним доступ. Примером таких вещей является прикладное ПО и мультимедийный контент. Наиболее распространенные методы для тестирования ПО:

1. Модульное тестирование. В данном методе с целью выявления возможных ошибок компоненты программы проверяются в отдельности. При этом необходимо хорошо знать состав программы, следовательно, этот вид тестирования осуществляется программистами, а не тестировщиками.

2. Интеграционное тестирование. В основном выявляет ошибки интерфейса. Компоненты программы, которые были проверены в ходе модульного тестирования, интегрируются друг с другом, а затем тестируются на наличие неполадок.

3. Системное тестирование. На данном этапе анализируется работа всей системы в целом при взаимодействии ее аппаратных и программных составляющих.

Основные виды тестов для проверки ПО:

1. Тестирование безопасности. Такой тип тестирования проверяет то, обеспечивает ли конкретное ПО конфиденциальность и функции аутентификации.

2. Стресс-тестирование. ПО подвергается таким условиям работы, которые выходят за рамки нормальных условий. При достижении некой критической точки результаты тестирования анализируются. Данный подход предназначен для выявления устойчивости ПО к нагрузкам.

3. Тестирование совместимости. В ходе теста ПО проверяется на совместимость с другими операционными системами, аппаратными платформами, web-браузерами и т.д.

4. Тестирование эффективности. Анализируется объем используемых аппаратных ресурсов при выполнении программой одной из операций.

5. Тестирование удобства пользования. Этот тест направлен на выявление степени удобства и практичности ПО для конечного пользователя.

3. Виды взаимодействия Интернет Вещей и возможные тестовые спецификации

Согласно рекомендации МСЭ У.2060 [9] возможны следующие виды взаимодействия различных типов Интернет Вещей (рисунок 2):

1. Обмен данными с использованием сети связи через шлюз;
2. Обмен данными с использованием сети связи без шлюза;
3. Обмен данными напрямую, т.е. без использования сети связи;
4. Сочетание первого и третьего сценариев;
5. Сочетание второго и третьего сценариев.

Следует отметить тот факт, что хотя на данной схеме показаны взаимодействия между объектами физического мира (обмен данными между устройствами), они также происходят и в информационном мире (взаимодействие виртуальных вещей), а также между физическим и информационным миром.

Далее приведены некоторые виды возможных тестов для Интернет Вещей, относительно используемых сетевых технологий:

1. Тестирование конфигурации сети, построенной по технологии Ethernet (стандарты 10Base-T, 100Base-T, 1000Base-T, 10GBase-T);

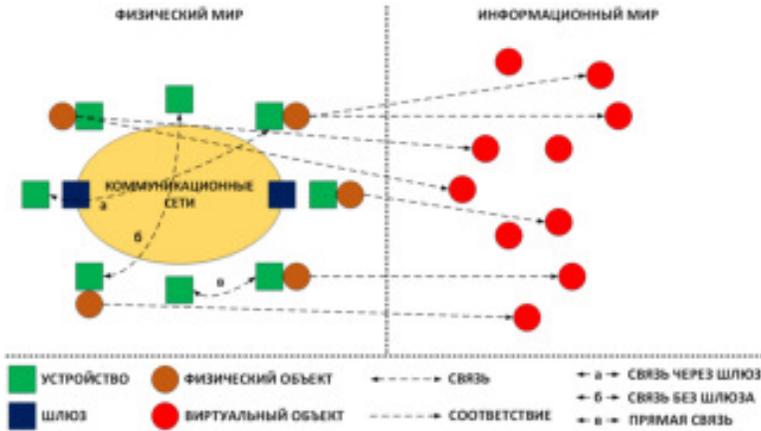


Рис. 2. Виды взаимодействия Интернет Вещей

2. Комплексное тестирование WLAN на базе стандарта IEEE 802.11 (Wi-Fi);

3. Определение коэффициента битовых ошибок при передаче данных по стандарту IEEE 802.15.1 (Bluetooth, Bluetooth LE);

4. Определение коэффициента ошибочных пакетов при передаче данных по стандарту IEEE 802.15.4 (ZigBee, 6LoWPAN);

5. Определение среднего значения скорости передачи данных к приемному терминалу при использовании технологий мобильной связи;

6. Тест с целью выбора прикладного протокола для Интернет Вещи относительно среднего времени отклика на запрос (RTT) при использовании протоколов MQTT, CoAP и HTTP.

4. Проведение натурального эксперимента

Для апробации предлагаемых тестовых спецификаций был проведен натуральный эксперимент, посвященный тестированию протоколов MQTT, CoAP и HTTP относительно времени отклика сервера на запрос. Следовательно, в ходе теста нужно отправлять запросы на сервер и отслеживать значения RTT, которые в дальнейшем необходимо проанализировать, и выявить наиболее подходящий протокол для использования в данной Интернет Вещи. Результаты проведенного эксперимента отражены на рисунке 3, который показывает, что наименьшее время отклика наблюдается при использовании протокола CoAP.

CoAP действительно является хорошим решением для межмашинных коммуникаций в Интернете Вещей. Данный протокол достаточно

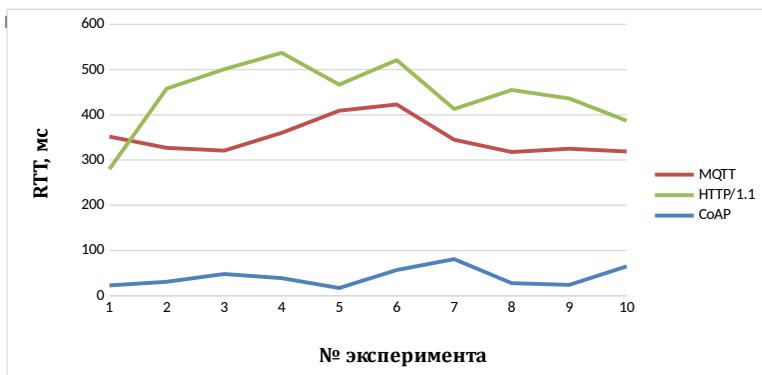


Рис. 3. Результаты проведенного эксперимента

легковесный и быстрый, к тому же имеет различные опции. Но к сожалению, клиентская часть в CoAP не может быть реализована при помощи обычных web-страниц, написанных на языке JavaScript, вследствие чего, приходится использовать дополнительные инструменты, например, Sorfer. Также программная реализация CoAP-сервера намного сложнее, чем у MQTT и HTTP.

5. Заключение

В статье были рассмотрены различные виды и методы тестирования устройств Интернета Вещей. Данные методы и тестовые спецификации могут быть использованы при проведении тестирования Интернет Вещей для обеспечения их стабильной работы. На базе одной из тестовых спецификаций проведен натурный эксперимент и представлены его результаты.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта N15-07-09431a "Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей".

Литература

1. Y.2069. Terms and definitions for the Internet of things // Recommendation ITU-T // Jul 2012.

2. *Kirichek R., Koucheryavy A.* Internet of Things Laboratory Test Bed // Lecture Notes in Electrical Engineering. 2016. Vol. 348. PP. 485-494.
3. *Кулик В.А., Киричек Р.В., Кучерявый А.Е.* Кулик В.А., Киричек Р.В., Кучерявый А.Е. Программно-аппаратный комплекс для тестирования устройств Интернета Вещей // Информационные технологии и телекоммуникации. 2015. № 4 (12). С. 67-76.
4. *Q.3900.* Methods of testing and model network architecture for NGN technical means testing as applied to public telecommunication networks // Recommendation ITU-T // September 2006.
5. *Васильев А.Б., Тарасов Д.В., Андреев Д.В., Кучерявый А.Е.* Тестирование сетей связи следующего поколения. М.:ФГУП ЦНИИС, 2008 г. – 144 с.
6. *Кулик В.А., Киричек Р.В., Бондарев А.Н.* Методы исследования беспроводных каналов связи Интернета Вещей в условиях совместной работы // Информационные технологии и телекоммуникации. - 2015. - № 1. - С. 106-114.
7. *Кучерявый А.Е.* Интернет Вещей // Электросвязь. - 2013. - № 1. с.21-24.
8. *Кучерявый А.Е., Прокопьев А.Е., Кучерявый Е.А.* Самоорганизующиеся сети. – СПб.: Любавич, 2011. – 312 с.
9. *Y.2060.* Overview of the Internet of things // Recommendation ITU-T // June 2012.

UDC 004.7

An Overview of Possible Testing Types and Methods for the Internet of Things

R. A. Dolgushev, R. V. Kirichek

*Department of Communication Networks and Data Transmission
The Bonch-Bruевич Saint-Petersburg State University of Telecommunications
pr. Bol'shevnikov 22, Saint-Petersburg, 193232, Russia*

This paper is devoted to the study of existing testing types and methods, that are used in communication networks, and to the selection of test specifications applicable in the field of the Internet of Things, which allow to estimate the various criteria, associated with the exchange of data between things, using network resources, protocols, etc. In addition, the full-scale experiment has been made in order to select an application protocol for the Internet thing concerning the definition of the average response time to a request (RTT) using such protocols as MQTT, COAP and HTTP.

Keywords: Internet Thing, IoT, FUSN, Ethernet, Wi-Fi, Bluetooth, Zig-Bee, MQTT, CoAP, HTTP.

УДК 514.7:514.8

Спинорное представление уравнений Максвелла

Е. Г. Ефери́на*, О. В. Кузнецова*, А. В. Королькова*,
Д. С. Кулябов*[†], Л. А. Севастьянов*[‡]

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

[†] *Лаборатория информационных технологий,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980*

[‡] *Лаборатория теоретической физики,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980*

Аннотация. Предпосылки Спиноры являются более специализированными объектами, чем тензоры. Поэтому обладают большим количеством свойств, нежели более общие объекты, такие как тензоры. Группа лоренцевых 2-спиноров является накрывающей группой группы Лоренца. Цель. Поскольку группа Лоренца является группой симметрии уравнений Максвелла, то предполагается оправданным использовать при записи уравнений Максвелла спиноры вместо тензоров. МЕТОДЫ Уравнения Максвелла записываются в форме лоренцевых спиноров. Также используется удобное представление лоренцевых спиноров через комплексные векторы Зильберштейна. РЕЗУЛЬТАТЫ В спинорном формализме (в представлении лоренцевых спиноров и векторов Зильберштейна) построен гамильтониан максвелловской оптики. При спинорной записи уравнения Максвелла приобретают вид, подобный уравнениям Дирака. Выводы При записи уравнений Максвелла в диракоподобном виде представляется возможным расширить инструментарий исследования за счёт методов квантовой теории поля. В этом виде наглядно представляется связь между гамильтонианами геометрической, параксиальной и волновой оптики.

Ключевые слова: спиноры, уравнения Максвелла, представление Зильберштейна.

1. Введение

Уравнения Максвелла имеют большое число представлений [1]. Принцип их введения следующий: каждое представление должно упрощать конкретное теоретическое или практическое исследование. К нашим интересам при исследовании уравнений Максвелла относятся их геометризация и гамильтонов формализм. В данной работе мы предлагаем на основе спинорного описания уравнений Максвелла [2–4] сконструировать дираковски-подобный гамильтониан. Предполагается, что данная форма позволит применять аппарат, разработанный в рамках квантовой теории, к изучению уравнений Максвелла.

Структура статьи следующая. В разделе 2 введены основные обозначения и соглашения. В разделе 3 даётся краткое описание уравнений Максвелла. В разделе 4 Вводится комплексное представление уравнений Максвелла. Далее, в разделе 5 даётся спинорное представление уравнений Максвелла. И в разделе 6, на основе комбинации результатов двух предыдущих разделов, получен дирако-подобный гамильтониан уравнений Максвелла.

2. Обозначения и соглашения

1. В работе используется нотация абстрактных индексов [5]. В данной нотации тензор как целостный объект обозначается просто индексом (например, x^i), компоненты обозначаются подчёркнутым индексом (например, $x^{\underline{i}}$).
2. Будем придерживаться следующих соглашений. Греческие индексы (α, β) будут относиться к четырёхмерному пространству и в компонентном виде будут иметь следующие значения: $\alpha = \overline{0, 3}$. Латинские индексы из середины алфавита (i, j, k) будут относиться к трёхмерному пространству и в компонентном виде будут иметь следующие значения: $\underline{i} = \overline{1, 3}$.
3. Запятой в индексе обозначается частная производная по соответствующей координате ($f_{,i} := \partial_i f$); точкой с запятой — ковариантная производная ($f_{;i} := \nabla_i f$).
4. Для записи уравнений электродинамики в работе используется система СГС симметричная.

3. Уравнения Максвелла

Уравнения Максвелла в 3-х мерной форме имеют вид:

$$\begin{aligned} \nabla_0 B^i &= -e^{ijk} \nabla_j E_k; \\ \nabla_i D^i &= 4\pi\rho; \\ \nabla_0 D^i &= e^{ijk} \nabla_j H_k - \frac{4\pi}{c} j^i; \\ \nabla_i B^i &= 0. \end{aligned} \tag{1}$$

где e^{ijk} — альтернирующий тензор, выражающийся через символ Леви-Чивиты $\varepsilon^{\underline{ijk}}$:

$$e_{\underline{ijk}} = \sqrt{3g} \varepsilon_{\underline{ijk}}, \quad e^{ijk} = \frac{1}{\sqrt{3g}} \varepsilon^{ijk}.$$

Запишем (1) через тензоры электромагнитного поля $F_{\alpha\beta}$ и $G_{\alpha\beta}$ [6]:

$$\begin{aligned}\nabla_\alpha F_{\beta\gamma} + \nabla_\beta F_{\gamma\alpha} + \nabla_\gamma F_{\alpha\beta} &= F_{[\alpha\beta;\gamma]} = 0, \\ \nabla_\alpha G^{\alpha\beta} &= \frac{4\pi}{c} j^\beta,\end{aligned}\quad (2)$$

где тензоры $F_{\alpha\beta}$, $F^{\alpha\beta}$, $G^{\alpha\beta}$ и $G_{\alpha\beta}$ имеют следующие компоненты

$$F_{\underline{\alpha\beta}} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B^3 & B^2 \\ -E_2 & B^3 & 0 & -B^1 \\ -E_3 & -B^2 & B^1 & 0 \end{pmatrix}, \quad G^{\underline{\alpha\beta}} = \begin{pmatrix} 0 & -D^1 & -D^2 & -D^3 \\ D^1 & 0 & -H_3 & H_2 \\ D^2 & H_3 & 0 & -H_1 \\ D^3 & -H_2 & H_1 & 0 \end{pmatrix},$$

E^i , H^i , $\underline{i} = \overline{1,3}$, — компоненты векторов напряжённости электрического и магнитного полей соответственно; D_i , B_i , $\underline{i} = \overline{1,3}$, — компоненты векторов электрической и магнитной индукции соответственно¹.

4. Комплексное представление уравнений Максвелла

Комплексное представление уравнений Максвелла рассматривалось разными авторами [7–9].

4.1. Общее представление

Зададим соответствие упорядоченной пары и комплексного 3-вектора

$$\begin{aligned}F^{\underline{i}} &\sim (E^i, B^i), & F^i &= E^i + iB^i; \\ G^{\underline{i}} &\sim (D^i, H^i), & G^i &= D^i + iH^i.\end{aligned}$$

Выразим напряжённость и индукцию через соответствующие комплексные векторы

$$\begin{aligned}E^i &= \frac{F^i + \bar{F}^i}{2}, & B^i &= \frac{F^i - \bar{F}^i}{2i}, \\ D^i &= \frac{G^i + \bar{G}^i}{2}, & H^i &= \frac{G^i - \bar{G}^i}{2i}.\end{aligned}$$

Введём два дополнительных комплексных вектора

$$K^i = \frac{G^i + F^i}{2}, \quad L^i = \frac{\bar{G}^i - \bar{F}^i}{2}. \quad (3)$$

¹Следует заметить, что именно B^i имеет физический смысл напряжённости магнитного поля.

Тогда уравнения (1) примут вид

$$\begin{aligned}\nabla_i(K^i + L^i) &= 4\pi\rho; \\ -i\nabla_0(K^i - L^i) + e^{ijk}\nabla_j(K_k - L_k) &= i\frac{4\pi}{c}j^i.\end{aligned}\quad (4)$$

4.2. Комплексное представление уравнений Максвелла в вакууме

Из соотношений $D^i = E^i$, $H^i = B^i$ и (3) получаем

$$K^i = E^i + iB^i = F^i, \quad L^i = 0.$$

Тогда уравнения (4) будут иметь вид

$$\begin{aligned}\nabla_i F^i &= 4\pi\rho; \\ -i\nabla_0 F^i + e^{ijk}\nabla_j F_k &= i\frac{4\pi}{c}j^i.\end{aligned}\quad (5)$$

4.3. Комплексное представление уравнений Максвелла в однородной изотропной среде

В однородной изотропной среде справедливы следующие соотношения $D^i = \varepsilon E^i$, $H^i = B^i$, где ε и μ — диэлектрическая и магнитная проницаемости.

Для упрощения получаемых выражений возможен следующий трюк. В (5) делаем формальную замену $c \rightarrow c' = \frac{c}{\sqrt{\varepsilon\mu}}$ (то есть заменяем скорость света в вакууме на скорость света в среде) и $j^\alpha \rightarrow \frac{j^\alpha}{\sqrt{\varepsilon}}$. Тогда получим

$$F^i = \sqrt{\varepsilon}E^i + i\frac{1}{\sqrt{\mu}}B^i.\quad (6)$$

Тогда уравнения Максвелла (1) приобретают вид:

$$\begin{aligned}\nabla_i F^i &= \frac{4\pi}{\sqrt{\varepsilon}}\rho; \\ i\frac{\partial F^i}{\partial t} &= \frac{c}{\sqrt{\varepsilon\mu}}e^{ijk}\nabla_j F_k - i\frac{4\pi}{\sqrt{\varepsilon}}j^i.\end{aligned}\quad (7)$$

Данное представление уравнений Максвелла имеет несколько наименований. В частности, оно известно как представление Зильберштейна [7, 8].

5. Спинорная запись уравнений Максвелла

Тензор электромагнитного поля $F_{\alpha\beta}$ и его компоненты $F_{\alpha\beta}$, $\underline{\alpha}$, $\underline{\beta} = \overline{0, 3}$, можно рассматривать в спинорной форме [5] (аналогично и для $G_{\alpha\beta}$):

$$\begin{aligned} F_{\alpha\beta} &= F_{AA'BB'}; \\ F_{\alpha\beta} &= F_{\underline{A}\underline{A}'\underline{B}\underline{B}'} g_{\underline{\alpha}}^{\underline{A}\underline{A}'} g_{\underline{\beta}}^{\underline{B}\underline{B}'}, \\ \underline{A}, \underline{A}', \underline{B}, \underline{B}' &= \overline{0, 1}, \quad \underline{\alpha}, \underline{\beta} = \overline{0, 3}, \end{aligned}$$

где $g_{\underline{\alpha}}^{\underline{A}\underline{A}'}$, $\underline{\alpha} = \overline{0, 3}$, — символы Инфельда–ван дер Вердена, определяемые в действительном спинорном базисе $\varepsilon_{\underline{A}\underline{B}}$ следующим образом [5]:

$$g_{\underline{\alpha}}^{\underline{A}\underline{A}'} := g_{\underline{\alpha}}^{\alpha} \varepsilon_{\underline{A}}^{\underline{A}} \varepsilon_{\underline{A}'}^{\underline{A}'}, \quad g_{\underline{A}\underline{A}'}^{\alpha} := g^{\alpha} \varepsilon^{\underline{A}} \varepsilon^{\underline{A}'}, \quad (8)$$

$$\varepsilon_{\underline{A}\underline{B}} = \varepsilon_{\underline{A}'\underline{B}'} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \varepsilon_{\underline{A}}^{\underline{A}} \varepsilon_{\underline{A}'}^{\underline{A}'} = \varepsilon_{\underline{A}}^{\underline{A}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (9)$$

Запишем уравнения Максвелла через спиноры.

Поскольку тензор $F_{\alpha\beta}$ действителен и антисимметричен, то его можно представить в виде

$$\begin{aligned} F_{\alpha\beta} &= \varphi_{AB} \varepsilon_{A'B'} + \varepsilon_{AB} \bar{\varphi}_{A'B'}, \\ *F^{\alpha\beta} &= -i\varphi^{AB} \varepsilon^{A'B'} + i\varepsilon^{AB} \bar{\varphi}^{A'B'}. \end{aligned} \quad (10)$$

здесь φ_{AB} — спинор электромагнитного поля:

$$\varphi_{AB} := \frac{1}{2} F_{ABC'} C' = \frac{1}{2} F_{AA'BB'} \varepsilon^{A'B'} = \frac{1}{2} F_{\alpha\beta} \varepsilon^{A'B'}. \quad (11)$$

Аналогично можно записать

$$\begin{aligned} G^{\alpha\beta} &= \gamma^{AB} \varepsilon^{A'B'} + \varepsilon^{AB} \bar{\gamma}^{A'B'}, \\ *G_{\alpha\beta} &= -i\gamma_{AB} \varepsilon_{A'B'} + i\varepsilon_{AB} \bar{\gamma}_{A'B'}. \end{aligned} \quad (12)$$

Заменяя в уравнении (2) абстрактные индексы α на AA' и β на BB' , запишем:

$$\nabla_{AA'} G^{AA'BB'} = \frac{4\pi}{c} j^{BB'}.$$

Используя соотношение (12) получим

$$\nabla^{AB'} \gamma_A^B + \nabla^{BA'} \gamma_{A'}^{B'} = \frac{4\pi}{c} j^{BB'}.$$

Аналогично, из (10) получим

$$\nabla^{A'B} \varphi_B^A - \nabla^{AB'} \bar{\varphi}_{B'}^{A'} = 0.$$

Таким образом полная система уравнений Максвелла в спинорном представлении имеет вид

$$\begin{aligned} \nabla^{A'B} \varphi_B^A - \nabla^{AB'} \bar{\varphi}_{B'}^{A'} &= 0, \\ \nabla^{AB'} \gamma_A^B + \nabla^{BA'} \gamma_{A'}^{B'} &= \frac{4\pi}{c} j^{BB'}. \end{aligned}$$

Система уравнений Максвелла в вакууме в спинорной форме запишется в виде одного уравнения [5]:

$$\nabla^{AB'} \varphi_A^B = \frac{2\pi}{c} j^{BB'}.$$

Выпишем компоненты спинора электромагнитного поля:

$$\begin{aligned} \varphi_{\underline{A}\underline{B}} &= \frac{1}{2} F_{\alpha\beta} \varepsilon^{A'B'} g_{\underline{A}A'} g_{\underline{B}B'}, \\ \underline{A}, \underline{A}', \underline{B}, \underline{B}' &= \overline{0, 1}, \quad \underline{\alpha}, \underline{\beta} = \overline{0, 3}. \end{aligned}$$

Используя (8), (9) и обозначив $F_i = E_i - iB_i$, можно записать:

$$\begin{aligned} \varphi_{00} &= \frac{1}{2} (F_1 - iF_2), \quad \varphi_{01} = \varphi_{10} = -\frac{1}{2} F_3, \\ \varphi_{11} &= -\frac{1}{2} (F_1 + iF_2). \end{aligned}$$

6. Дирако-подобный гамильтониан

Из спинора второй валентности φ_{AB} (11) сконструируем спинор первой валентности ψ^a . При этом применим трюк (6).

$$\psi^a = \begin{pmatrix} \psi^1 \\ \psi^2 \\ \psi^3 \\ \psi^4 \end{pmatrix} = 2 \begin{pmatrix} \varphi^{00} \\ \varphi^{01} \\ \varphi^{10} \\ \varphi^{11} \end{pmatrix} = \begin{pmatrix} -F^1 + iF^2 \\ F^3 \\ F^3 \\ F^1 + iF^2 \end{pmatrix}.$$

Соответственно ток принимает вид:

$$\xi^a = \begin{pmatrix} -j^1 + ij^2 \\ j^3 - \sqrt{\varepsilon} \rho \\ j^3 + \sqrt{\varepsilon} \rho \\ j^1 + ij^2 \end{pmatrix}.$$

Тогда система уравнений (7) приобретает следующий дирако-подобный вид:

$$\frac{\partial \psi^a}{\partial t} = -\frac{c}{\sqrt{\varepsilon \mu}} e^{ijk} \nabla_j \gamma_k^a{}_b \psi^b - \frac{4\pi}{\sqrt{\varepsilon}} \xi^a.$$

Из структуры уравнения Дирака

$$i\hbar \frac{d\psi}{dt} = [c\gamma^i p_i + \gamma^0 mc^2] \psi = H\psi$$

можно записать дирако-подобный гамильтониан уравнений Максвелла при отсутствии токов:

$$H = -\frac{c}{\sqrt{\varepsilon \mu}} e^{ijk} \nabla_j \gamma_k^a{}_b.$$

7. Заключение

В статье представлен способ записи уравнений Максвелла в спинорном представлении и записан гамильтониан уравнений Максвелла в диракоподобном виде. Представляется, что данное представление может быть интересно в плане применения методов квантовой теории к изучению электромагнитных явлений.

Благодарности

Работа частично поддержана грантами РФФИ № 14-01-00628, 15-07-08795 и 16-07-00556.

Литература

1. Kulyabov D. S., Korolkova A. V., Korolkov V. I. Maxwell's Equations in Arbitrary Coordinate System // Bulletin of Peoples' Friendship University of Russia. Series "Mathematics. Information Sciences. Physics". — 2012. — no. 1. — P. 96–106. — arXiv : 1211.6590.
2. Laporte O., Uhlenbeck G. E. Application of spinor analysis to the Maxwell and Dirac equations // Physical Review. — 1931. — Vol. 37, no. 11. — P. 1380–1397.
3. Aste A. Complex Representation Theory of the Electromagnetic Field // Journal of Geometry and Symmetry in Physics. — 2012. — no. 28. — P. 47–58. — arXiv:1211.1218v3.
4. Bialynicki-Birula I. Photon Wave Function // Progress in Optics. — Elsevier, 1996. — Vol. 36. — P. 245–294. — 0508202.
5. Пенроуз Р., Риндлер В. Спиноры и пространство-время. Два-спинорное исчисление и релятивистские поля. — М. : Мир, 1987. — Т. 1.

6. Minkowski H. Die Grundlagen für die electromagnetischen Vorgänge in bewegten Körpern // *Math. Ann.* — 1910. — H. 68. — S. 472–525.
7. Silberstein L. Elektromagnetische Grundgleichungen in bivectorieller Behandlung // *Annalen der Physik.* — 1907. — Vol. 327, no. 3. — P. 579–586.
8. Silberstein L. Nachtrag zur Abhandlung über „Elektromagnetische Grundgleichungen in bivectorieller Behandlung“ // *Annalen der Physik.* — 1907. — Vol. 329, no. 14. — P. 783–784.
9. Стрэттон Д. А. Теория электромагнетизма. — М.-Л.: ГИТТЛ, 1948.

UDC 514.7:514.8

Spinor representation of Maxwell's equations

E. G. Eferina*, O. V. Kuznetsova*, A. V. Korolkova*,
D. S. Kulyabov*[†], L. A. Sevastianov*[‡]

* *Department of Applied Probability and Informatics
RUDN University
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

[†] *Laboratory of Information Technologies
Joint Institute for Nuclear Research
Joliot-Curie 6, Dubna, Moscow region, 141980, Russia*

[‡] *Bogoliubov Laboratory of Theoretical Physics
Joint Institute for Nuclear Research
Joliot-Curie 6, Dubna, Moscow region, 141980, Russia*

BACKGROUND Spinors are more special objects than tensor. Therefore possess more properties than the more generic objects such as tensors. Group of Lorentz two-spinors is the covering group of the Lorentz group. **PURPOSE** Since the Lorentz group is a symmetry group of Maxwell's equations, it is assumed to reasonable to use when writing the Maxwell equations Lorentz two-spinors and not tensors. **METHOD** We write the Maxwell's equations using Lorentz two-spinors. Also used a convenient representation of Lorentz two-spinors in terms of the Riemann-Silberstein's complex vectors. **RESULTS** In the spinor formalism (in the representation of the Lorentz spinors and Riemann-Silberstein's vectors) we have constructed the Hamiltonian of Maxwellian optics. With the use of spinors Maxwell's equations take the form similar to the Dirac equation. **CONCLUSIONS** For Maxwell's equations in the Dirac-like form we can expand research methods at the expense of the methods of quantum field theory. In this form, clearly visible the connection between the Hamiltonians of geometric, beam and Maxwellian optics.

Keywords: spinors, Maxwell's equations, Riemann-Silberstein's vectors.

УДК 681.513.7

Распознавание точек привязки трехмерных объектов по стереоизображению в системах технического зрения

М. Б. Фомин*, А. Е. Иванов*

** Кафедра информационных технологий,
Российский университет дружбы народов,
ул. Миклухо-Макляя, д.6, Москва, Россия, 117198*

Аннотация. В процессе реконструкции формы трехмерного тела по стереоизображению возникает задача распознавания точек привязки на поверхности объекта. Эти точки используются в алгоритме триангуляции, когда формируется описание модели объекта. Один из способов уточнения координат точек привязки – анализ карты диспаратета для выделенного сегмента стереоизображения. Предлагаемый метод распознавания основан на использовании марковской сети для описания карты диспаратета, преобразовании графа этой сети путём анализа распределения яркостей пикселей изображения. Возможность выбора приоритетного направления смещения при обходе графа карты диспаратета обеспечивает эффективность работы алгоритма, основанного на описанном методе.

Ключевые слова: стереоизображение, эпиполярная геометрия, карта диспаратета.

1. Введение

Задача реконструкции трехмерного описания по информации, предоставляемой системой технического зрения, возникает при построении рельефа местности, задании формы трехмерного тела, определении местоположения и ориентации объекта. При её решении может быть использована пара изображений, сделанных стереокамерой – откалиброванной системой двух камер с известными преобразующими характеристиками. Знание этих характеристик позволяет выполнить ректификацию изображений – сопоставление строк, соответствующих одной горизонтальной линии на сцене [1]. Такая предварительная подготовка позволяет перейти к анализу изображений и выделить соответствующие друг другу пары точек.

2. Постановка задачи

После выполнения ректификации появляется возможность построения сравнения левого и правого изображений с целью выявления связанных пикселей. Пара таких пикселей, принадлежащих разным изображениям, соответствует одной и той же точке на поверхности объекта сцены. Решить задачу распознавания связанных пикселей можно путем анализа распределения яркости в сопоставляемых

строках изображений, выявлением цветовых сегментов и установлением соответствия между границами цветовых сегментов, найденных на левом и правом изображениях. В процессе сопоставления последовательно фиксируются пиксели в строке левого изображения, и для каждого выбранного пикселя выделяется диапазон в строке правого изображения, который содержит пиксель, соответствующий зафиксированному. Распознать в диапазоне правого изображения этот пиксель можно путем анализа статистических характеристик распределения яркости в левой и правой строках в случае, если эти характеристики близки.

При анализе распределения яркости в окрестности выбранного пикселя левого изображения и в выделенном диапазоне правого изображения может возникать ситуация неоднозначного выбора пары. Возможны разные причины возникновения такой ситуации. Пикселу на одном изображении может соответствовать несколько пикселей на другом вследствие различия ракурсов. Элементы сцены, имеющиеся на одном изображении, могут отсутствовать на другом из-за эффекта загораживания. Наконец, наиболее частой причиной является монотонность текстуры изображенных объектов, затрудняющая процесс сопоставления. В таких ситуациях уточнение варианта выбора пары пикселей можно произвести путем замены используемых показателей корреляции яркостей или путем анализа яркостных характеристик в строках, лежащих выше и ниже анализируемой строки.

Задача распознавания пар связанных пикселей решается успешно при условии, что пиксели лежат на границах цветовых сегментов стереоизображения. В результате сопоставления на стереоизображении выделяется набор пар пикселей, для которых может быть произведена процедура триангуляции. Она позволяет определить глубину по координате Z для соответствующих пикселям точек сцены. Координаты этих пикселей могут быть использованы для задания точек привязки при моделировании поверхностей объектов сцены [2]. Кроме того, параметры этих пикселей могут быть использованы в задаче экстраполяции при построении полной карты диспаратета, которая формируется на базе левого изображения, и в которой для каждого пикселя задается дополнительный параметр – Z -координата соответствующей ему точки сцены. Одновременно для каждого пикселя p задается значение сдвига ξ – изменение для координаты X пикселя, которое необходимо учесть, чтобы получить значение координаты X для пикселя, принадлежащего правому изображению и соответствующего пикселю карты диспаратета:

$$\xi_p = X_{pr} - X_{pl},$$

где $pl = p$ – пиксель на карте диспаратета (он же – пиксель левого изображения), pr – соответствующий ему пиксель правого изображения.

Для всех пикселей сдвиг имеет отрицательное значение или равен нулю (в случае «очень большого» удаления соответствующей ему точки сцены). Для пикселей, параметр Z для которых получен экстраполяцией, значение сдвига может быть нецелым. В этом случае оно должно быть округлено до ближайшего целого значения.

Алгоритм распознавания границ цветowych сегментов относится к задачам распознавания, существенно зависит от особенностей изображения и характеризуется низкой точностью. В следствие этого построенная карта диспаратета, как правило, описывает сцену с погрешностью, которая выражается в резких перепадах по глубине на поверхностях объектов. Для адекватного описания сцены требуется произвести уточнение карты диспаратета.

3. Уточнение карты диспаратета

Карта диспаратета, полученная при анализе стереоизображения в результате построчечного сравнения, содержит погрешности, вызванные неточным распознаванием границ цветowych сегментов (результат – резкие перепады значения Z по горизонтали карты) и независимым анализом строк изображения (результат – «лестница» по вертикали карты). Для удаления с карты диспаратета описанных погрешностей может быть использован глобальный алгоритм, позволяющий добиться выравнивания значений координаты Z в областях изображения, соответствующих поверхностям объектов сцены, без усреднения по глубине в тех местах изображения, в которых резкое изменение координаты Z обосновано – в местах перехода с одного объекта на другой, расположенный на отличающейся глубине. Для количественного анализа «не гладкости» карты диспаратета можно по аналогии с подходом, используемым в термодинамике, ввести функцию «энергии», значение которой увеличивается при наличии на карте «возбуждений» по параметру Z . Выравнивание можно производить с использованием принципа минимизации функции «энергии». Построим для карты диспаратета граф $G = (P, E)$, узлами которого являются пиксели $p \in P$, а ребрами $e \in E$ – связи между ближайшими пикселями. При вычислении «энергии» карты будем использовать формулу

$$E(G) = E_{data}(G) + E_{smooth}(G),$$

где

$$E_{data}(G) = \sum_{p \in P} m(p),$$

– унарный потенциал, являющийся мерой цветового сходства,

$$m(p) = |I_{pr} - I_{pl}|,$$

I_{pr} и I_{pl} – значения яркости пикселя p соответственно на правом и левом изображениях,

$$E_{smooth}(G) = \sum_{e \in E} S(e)$$

– парный потенциал, являющийся мерой «не гладкости», заданный через штрафную функцию $S(e)$, которую можно определить с использованием значений Z -координат пикселей p и q на карте диспаратета, определяющих ребро e . Штрафная функция $S(e)$ может быть задана разными способами:

$$S(e) = Const * |Z_p - Z_q|,$$

или

$$S(e) = \begin{cases} 0, & |Z_p - Z_q| < M \\ Const, & |Z_p - Z_q| \geq M, \end{cases}$$

M – некоторое пороговое значение.

Второй вариант задания функции $S(e)$ в процессе выравнивания карты диспаратета приводит к меньшему усреднению параметра Z в тех местах изображения, в которых имеется переход с одного объекта сцены на другой.

Граф карты диспаратета с заданной на нем функцией энергии можно рассматривать как марковскую сеть. При этом ребра графа интерпретируются как связи между пикселями, описывающие зависимости между значениями параметра Z этих пикселей, которые необходимо учитывать при выравнивании карты диспаратета.

Выполнение глобального алгоритма состоит в многократном обходе графа карты диспаратета. На каждом шаге обхода анализируется один из узлов p графа. При этом сравниваются ξ_p – значение сдвига для этого узла и ξ_p^* – среднее значение сдвига по пикселям, смежным с пикселем p . В случае если $|\xi_p^* - \xi_p| < L_1$, где L_1 – некоторое пороговое значение, для пикселя p значение смещения ξ_p меняется на величину, приближающую его к ξ_p^* . Это изменение выбирается таким образом, чтобы задаваемая им вариация соответствовала локальному минимуму энергии. Одновременно при выполнении условия $|\xi_{p^*} - \xi_p| < L_2$, где p^* – узел, смежный с пикселем p , а L_2 – некоторое пороговое значение, из графа карты диспаратета может удаляться ребро, соединяющие p и p^* . Могут быть удалены одно вертикальное и одно горизонтальное ребро, при этом приоритет отдается ребрам, для которых модуль разности значений сдвигов меньше. После обхода графа производится перерасчет значений параметра Z карты диспаратета в соответствии с уточненными значениями сдвига для узлов графа.

Описанная процедура обхода графа карты диспаратета выполняется многократно. В качестве условия прекращения этого процесса можно использовать условие выполнения неравенства

$$E_i(G) - E_{i+1}(G) < \varepsilon,$$

где i и $i + 1$ – номера итераций, ε – малый параметр.

Глобальный алгоритм допускает распараллеливание выполнения. Для распараллеливания необходимо разбить граф на части, и решить задачу минимизации энергии для каждого из подграфов.

Другое возможное направление оптимизации глобального алгоритма – ограничение набора пикселей, анализируемых при обходе карты диспаратета. Действительно, с высокой вероятностью необходимость перестройки карты диспаратета возникает лишь для пикселей, лежащих на границе цветовых сегментов. Можно ограничиться анализом только этих пикселей.

Структура алгоритма такова, что используемые в нем вычисления могут быть выполнены с помощью библиотеки DirectX на видеокarte. Поскольку алгоритмы используют только простые алгебраические вычисления, эта возможность относится как к задаче построения карты диспаратета, так и к глобальному алгоритму, уточняющему карту диспаратета. В случае если ресурсы библиотеки DirectX реализованы на видеокarte, работа алгоритма может быть значительно ускорена, в том числе и за счет разбиения графа карты диспаратета на части и выполнения алгоритма для разных частей в разных потоках распараллеленных вычислений.

4. Заключение

В работе были рассмотрены варианты улучшения способа построения карты диспаратета для стереоизображения путем применения модифицированного глобального алгоритма, в котором граф карты диспаратета, используемый при преобразовании значений координаты Z для описания возможных сдвигов пикселей, упрощается за счет удаления части ребер, а анализ яркости пикселей производится только на границах выявленных цветовых сегментов. Алгоритм может выполняться на видеокarte с использованием библиотеки DirectX в режиме распараллеливания. При этом в случае управления беспилотным средством передвижения, оснащенным системой технического зрения, может быть обеспечена скорость, достаточная для восстановления в режиме on-line параметров трехмерного описания рельефа местности.

Литература

1. *Hartley R., Zisserman A. Multiple View Geometry in Computer Vision.* — Cambridge University Press, 2004. — ISBN: 0-521-54051-8.

2. Калиниченко А.В., Свешникова Н.В., Юрин Д.В. Эпиполярная геометрия и оценка ее достоверности по результатам восстановления трехмерной сцены алгоритмами факторизации [Электронный ресурс] // International Conference Graphicon, 2006. – Режим доступа: http://graphicon.ru/html/2006/wr22_85_KalinichenkoSveshnikovaYurin.pdf.

UDC 681.513.7

Recognition of anchor points on three-dimensional objects by stereo images in machine vision systems

М. В. Fomin*, А. Е. Ivanov*

** Department of Information Technologies
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

The problem of anchor points' recognition on the object surface occurs in the process of reconstruction of the three-dimensional shape of the body from the stereo image. These points are used in a triangulation algorithm that generates description of the object model. One way to recognize reference points – analysis of the disparity map for the selected segment of the stereo image. The proposed detection method is based on Markov network that describes the disparity maps, transforming the graph of this network by analyzing the distribution of pixel brightness in corresponding image segment. Ability to select priority directions of displacement when traversing image segment provides the effectiveness of the algorithm, based on the described method.

Keywords: stereo image, epipolar geometry, disparity map.

УДК 004.7

Исследование и разработка методов управления узлом сенсорной сети на основе нейронных сетей

М. А. Герасименко

*Департамент Компьютерной инженерии Московского института
электроники и математики*

*Национального исследовательского университета Высшая школа
экономики (выпускница),*

ул. Таллинская, д.34, Москва, Россия, 123458

Аннотация. В работе представлены результаты исследования и разработки метода повышения энергоэффективности и качества управления узлом в беспроводных сенсорных сетях. Предлагается модель беспроводной сенсорной сети, учитывающая подзарядку от окружающей среды, метод распределения нагрузки на узлы беспроводной сенсорной сети на основе нейронных сетей. Управление сенсорной сетью сводится к задаче классификации по нескольким состояниям узла и тестируется на имитационной модели сенсорной сети, выполненной в Matlab.

Ключевые слова: сенсорные сети, имитационное моделирование, энергоэффективность, нейронные сети, протоколы для сенсорных сетей.

1. Введение

Статья посвящена актуальной теме – исследованию и разработке методов повышения энергоэффективности беспроводной сенсорной сети, далее БСС. БСС – сеть из маломощных передатчиков, которые оснащены датчиками для снятия характеристик окружающей среды, например, температуры, влажности, движения, радиоактивного излучения и других характеристик. В современном мире БСС используются как промышленные системы мониторинга, например, беспроводная система охраны труда на вредном производстве, а также беспроводная система мониторинга параметров машин и механизмов на стадии приемо-сдаточных испытаний [1].

Особенностями БСС являются малая мощность, ограниченный объем аккумулятора и внутренней памяти, поэтому исследование возможности подзарядки узла от окружающей среды и энергоэффективное управление узлом является актуальным.

Целью исследования является повышение эффективности метода управления узлом сенсорной сети, используя метод на основе нейронных сетей, для верификации гипотезы о возможности организации распределенных вычислений в БСС.

Для достижения поставленной цели выделяется ряд задач:

- Обзор и анализ существующих методов управления узлами БСС.
- Определение компонентов имитационной модели, и их свойств.

- Определение разработанного метода управления БСС на основе нейронных сетей и его особенностей.
- Тестирование результатов применения разработанного метода.

2. Основная часть

Узлы БСС только получают информацию от своих датчиков, но не обрабатывают ее, поэтому для того, чтобы устройство обработки (анализа) данных (сервер, базовая станция и другие аналоги) получило данные, необходим специальный протокол для БСС, стремящийся наиболее выгодно, с точки зрения минимизации потребляемой энергии, передать данные от узлов к базовой станции. Таких протоколов существует много, но принцип действия их различен. К примеру, протоколы для БСС можно разделить на протоколы прямого приближения, протоколы, ориентированные на местоположение узлов (в таком случае при маршрутизации учитывается только местоположение узлов, за счет GPS приемника малой мощности, встроенного в узлы), атрибутно ориентированные протоколы. [2]. Также, протоколы для БСС можно разделить на протоколы, сводящие сеть к древовидной структуре, и протоколы, разбивающие сеть на кластеры. В таблице 1 представлен сравнительный анализ некоторых протоколов.

Таблица 1

Сравнительный анализ протоколов БСС

	Тип	Эффективность	Ограничение	Узлы	Стабильность
LEACH	Кластеры	1	К-во головных узлов	Статические	1
SEER	Деревья	1	Нет	Статические	2
FEAR	Деревья	2	Нет	Статические	1
DECA	Кластеры	2	Нет	Статические	2

1 - среднее значение;

2 - высокое значение.

Протокол LEACH (Low Energy Adaptive Clustering Hierarchy) – один из первых энергоэффективных протоколов для сенсорных сетей, и на момент разработки давал существенно уменьшение потребления энергии. Протокол используется сейчас и имеет множество вариаций и доработок. LEACH используется для статических узлов БСС и контролирует топологию сети, постоянно разбивая ее на кластеры согласно

объему данных, который необходимо передать. Внутри кластера происходит выбор головного узла (cluster head). Это производится для предотвращения чрезмерного потребления энергии при передаче большого объема на большие расстояния каждым узлом самостоятельно. Протокол включает в себя агрегирование данных и сокращение рассеивания энергии в 8 раз по сравнению с существующими на момент создания [3].

LEACH был выбран как базовый, на основе принципов которого разрабатывается собственный протокол на основе нейронных сетей. LEACH показывает стабильные результаты, но нуждается в повышении энергоэффективности и подходит для БСС с большим количеством узлов, а также организации в большой сети распределенных вычислений. В таблице 2 приведено сравнение разработанного протокола с существующей версией LEACH.

$$T(n) = \begin{cases} \frac{P}{1-P*(r \bmod n)}, \\ 0, ?G \end{cases} \quad (1)$$

$$E = Nd(n).En - ((ETX + EDA) * (4000)Emp * 4000 * (dst^4)) \quad (2)$$

$$dst = \sqrt{((Nd(n).xd - Nd(n+1).xd)^2 + (Nd(n).yd - Nd(n+1).yd)^2)} \quad (3)$$

$$ETX = 5 * (10)^{-8}$$

$$ERX = 5 * (10)^{-8}$$

$$Efs = (10)^{-11}$$

$$Emp = 13 * (10)^{-16}$$

$$EDA = 5 * (10)^{-9}$$

r – текущий раунд, на котором заново выбирается головной узел;

P – вероятность выбора головным узлом, которая задается изначально

n – указатель на текущий узел.

$Nd(n).En$ – запас энергии узла n ;

dst – расстояние между узлами;

$Nd(n).xd$ – координата x узла n ;

$Nd(n).yd$ – координата y узла n .

Таким образом, ожидается повышение энергоэффективности благодаря использованию нового метода на основе нейронных сетей, а также модуля подзарядки от окружающей среды.

Таблица 2

Сравнение протокола LEACH с разрабатываемым

	LEACH	Разрабатываемый протокол
Статические узлы	Да	Да
Кластеры	Да	Нет жесткого разбиения на кластеры
Наличие головного узла	Да	Да
Способ выбора головного узла	Узел n , произвольно генерирует символ между 0 или 1, затем происходит расчет формулы (1). Если сгенерированный узлом символ меньше $T(n)$, то узел становится головным узлом.	Узел хранит программу на основе нейронных сетей, которая классифицирует текущее состояние узла по нескольким состояниям. Состояние узла изменяется с течением времени.
Подзарядка от окружающей среды	Нет	Да
Ограничение головных узлов	Да	Да
Расчет потребляемой энергии	Согласно формулам (2), (3).	Согласно формулам (2), (3).

Разработанная модель БСС для проведения тестов является имитационной, выполнена в среде Matlab и представляет собой совокупность типов узлов БСС, помещения, в которой они находятся, сценария их поведения. Типами узлов являются: узел-вычислитель, узел-передатчик, узел в подзарядке, узел в базовом состоянии. Типы узлов аналогичны физическим состояниям, в которых может находиться узел БСС: прием, передача, подзарядка, вычисление, ожидание (спящий режим). Основным исследуемым параметром модели является потребляемая энергия, которая зависит от текущего состояния.

Метод управления имитационной моделью основан на нейронных сетях – одном из методов автоматического управления, который

успешно решает задачи классификации, самообучается, а также прогнозирует результаты. Поскольку задача управления узлом БСС свелась к задаче классификации по нескольким состояниям, то метод нейронных сетей подойдет для решения этой задачи, а самообучение метода – уникальный способ повысить качество управления узлом БСС.

Нейронная сеть состоит минимум из одного слоя нейронов, у каждого из которых есть несколько входов, с некоторыми весами для каждого, и выходы. На вход подается либо внешний сигнал, либо выход предыдущего слоя нейронов. Нейрон – функция, сопоставляющая входы и выходы. Входы нейрона – закодированные числом свойства всех возможных состояний узла БСС, а выходы – число, соответствующее одному из 5 типов узлов. То есть, задачей нейронной сети является определение состояния узла БСС, в которое он должен перейти в зависимости от своих параметров, которые изменяются со временем. В качестве входов выбраны следующие атрибуты: флаг «Задача», флаг «Данные», запас энергии, приоритет, требуемая энергия.

Метод уже был протестирован, и были выявлены зависимости потребления энергии от числа узлов, плотности распределения узлов, поэтому необходимо повышать эффективность метода нейронных сетей. Критерием эффективности метода является процент узлов, которые были верно классифицированы по типам, в зависимости от их состояний.

Существует несколько вариаций использования нейронных сетей, поэтому необходимо выполнить сравнение. Для сравнения выбраны различные методы обучения нейронной сети: обучение с заранее определенными значениями на входах и выходах (с учителем) и обучение только с определенными входами (без учителя). Для метода без учителя не требуются заранее подготовленные выходные данные, а предполагается, что сеть обучится самостоятельно [4]. Среда Matlab, где проводится моделирование, предлагает тип нейронной сети «Самоорганизующиеся карты», которые обучаются классификации входных векторов по сходству и используются для задач классификации. «Карты» могут сохранять топологию входных векторов, назначая близлежащие входы к близлежащим категориям [5], а это подходит под поставленную задачу.

Сравнение различных типов нейронных сетей представлены в таблице 3. Существенным отличием между типами является отсутствие массива целей у нейронной сети «без учителя». Это, с одной стороны, упрощает процесс обучения, но с другой, повышает требования к качеству входных данных.

Нейронная сеть «без учителя» не смогла самостоятельно классифицировать заданную выборку на 5 классов, в отличие от нейронной сети «с учителем». В дальнейшем предполагается использовать нейронную сеть «с учителем», а также повышать качество обучающей выборки.

Сравнение методов нейронных сетей

Таблица 3

	С учителем	Без учителя
Входные данные	P, T	P
Вычисление выходного сигнала A	Да	Да
Сравнение A с T	Да	Нет
Вычисление ошибки и ее обратное распространение	Да	Нет
Условие завершения обучения	Если ошибка ниже заданного порога по всему массиву P.	Если на выходе согласованные векторы.
Корректировка весов	Да	Да
Итераций до подбора весов	6	100
Ошибка отклонения эпох обучения	10-5	Отсутствует
Эффективность метода,	18%	7%

$P = [i; j]$ – массив входных сигналов, где i – строки, являющиеся примерами возможных значений параметров типов узлов (в данном случае, $i = 5$), а j – столбцы, являющиеся самими параметрами (в данном случае, $j = 5$, так как решено использовать 22 тестовых примера для нейронной сети).

$T = [1; i]$ – массив целей, который показывает принадлежность совокупности параметров строки на шаге 1 к определенному типу узла.

3. Заключение

В работе представлен сравнительный анализ существующих протоколов управления узлами БСС, определены компоненты имитационной модели БСС и метод управления на основе нейронных сетей, также протестированы различные разновидности метода нейронных сетей. В результате тестирования нейронная сеть, обучающаяся «с учителем»

показала более высокие результаты, чем нейронная сеть, обучающаяся «без учителя». Низкие показатели эффективности метода могут быть причиной малого количества входных сигналов, а также низкого качества входных данных. Предполагаемым преимуществом предлагаемого метода управления БСС является повышение энергоэффективности сети настолько, чтобы можно было отказаться от кластерной архитектуры и протоколов, направленных на передачу данных от узла на сервер для последующей обработки. Вместо этого, часть данных будет обработана узлом БСС, поддерживающим распределенные вычисления.

Литература

1. *Баскаков С.* Беспроводные сенсорные сети на базе платформы Meshlogic. —Электронные компоненты. —№8. —Р. 361–363.—2008.
2. *Ting D., Haiping H., Yang L., Ruchuan W., Xinxing P.* Research on Migration Strategy of Nobile Agent in Wireless Sensor Network — International Journal of Distributed Sensor Networks/ —2013.
3. *Tao L.* Improvement for LEACH Algorithm in Wireless Sensor Network. — Industrial Electronics and Applications, the 5th IEEE Conference.— Р. 15–17.— 2010.
4. *Schmidhuber J.* Deep Learning in Neural Networks: An Overview . —The Swiss AI Lab IDSIA. —[Электронный ресурс]. — URL:<http://arxiv.org/pdf/1404.7828.pdf>. —(дата обращения 01.03.2016).
5. *Kaski S.* Deep Learning in Neural Networks: An Overview . — Acta Polytechnica Scandinavica. —[Электронный ресурс]. — URL:<http://users.ics.aalto.fi/sami/thesis/>. —(дата обращения 10.03.2016).

UDC 004.7

Neuron Networks

M. A. Gerasimenko

Department of Computer Engineering, HSE Moscow Institute of Electronics and Mathematics (MIEM HSE), Tallinskaya Ulitsa, 34, Moscow, Russia, 123458

This paper is about increase of energy efficiency of wireless sensor networks by development new protocol based on neural networks, energy harvesting and non-cluster architecture of wireless sensor networks.

Keywords: This paper is about increase of energy efficiency of wireless sensor networks by development new protocol based on neural networks, energy harvesting and non-cluster architecture of wireless sensor networks.

УДК 004.94+519.62

Реализация винеровского и пуассоновского стохастических процессов в OpenModelica

М. Н. Геворкян*, А. В. Демидова*, А. В. Королькова*,
Д. С. Кулябов*[†], Л. А. Севастьянов*[‡]

* Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклуто-Маклая, д.6, Москва, Россия, 117198

[†] Лаборатория информационных технологий,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980

[‡] Лаборатория теоретической физики,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980

Аннотация. В открытой реализации компонентно-ориентированного языка моделирования OpenModelica нет средств для моделирования стохастических процессов. В статье рассматривается проблема добавления генераторов случайных распределений, отличных от равномерного, в частности нормального и пуассоновского. Нормальное распределение затем используется для моделирования винеровского случайного процесса, который играет ключевую роль в стохастических дифференциальных уравнениях.

Ключевые слова: псевдослучайные числа, стохастические процессы, процесс Винера, Modelica.

1. Введение

В данной работе рассматривается вопрос генерации нормального, пуассоновского и экспоненциального распределений. Такой выбор мотивирован использованием этих распределений в теории стохастических дифференциальных уравнений. Наиболее общий вид таких уравнений использует два случайных процесса: винеровский и пуассоновский. Винеровский процесс позволяет учесть имплицитную стохастичность моделируемой системы, а пуассоновский процесс — внешнее воздействие. Кратко опишем структуру статьи.

2. Генерирование нормального и пуассоновского распределений

Генераторы псевдослучайных равномерно распределенных чисел являются основой для получения других псевдослучайных последовательностей. Однако, большинство алгоритмов требуют задания случайного числа из интервала $[0, 1]$, в то время как подавляющее большинство генераторов равномерно распределенных псевдослучайных чисел дают последовательность из интервала $[0, m]$, где число m зависит от алгоритма, разрядности операционной системы и процессора.

Для получения чисел из интервала $[0, 1]$ можно поступить двумя способами. Во первых можно нормировать полученную последовательность случайных чисел, поделив каждый ее элемент на максимальный элемент последовательности. Такой подход гарантированно даст 1 в качестве одного из случайных чисел. Однако такой способ плох, когда последовательность псевдослучайных чисел слишком велика и не умещается в оперативную память. В этом случае лучше использовать второй способ, а именно поделить каждое сгенерированное число на m .

2.1. Генерирование нормального распределения

Метод генерации нормально распределенных псевдослучайных величин предложен в 1958 году Дж. Э. Р. Боксом и П. Мюллером [1] и назван в их честь *преобразованием Бокса-Мюллера*. Метод основан на простом порообразовании, задаваемом двумя формулами. Данное преобразование обычно записывается в двух форматах:

- стандартная форма (она как раз и предложена авторами статьи [1]),
- полярная форма (предложена Дж. Беллом [2] и Р. Кнопом [3]).

Стандартная форма. Пусть x и y — два независимых, равномерно распределенных псевдослучайных числа из интервала $(0, 1)$, тогда числа z_1 и z_2 вычисляются по формуле

$$z_1 = \cos(2\pi y)\sqrt{-2 \ln x}, \quad z_2 = \sin(2\pi y)\sqrt{-2 \ln x}$$

и являются независимыми псевдослучайными числами, распределенными по стандартному нормальному закону $\mathcal{N}(0, 1)$ с математическим ожиданием $\mu = 0$ и стандартным среднеквадратичным отклонением $\sigma = 1$.

Полярная форма. Пусть x и y — два независимых, равномерно распределенных псевдослучайных числа из интервала $[-1, 1]$. Вычислим вспомогательную величину $s = x^2 + y^2$, если $s > 1$ и $s = 0$, то значения x и y следует отбросить и проверить следующую пару. Если же $0 < s \leq 1$ тогда числа z_1 и z_2 вычисляются по формуле

$$z_1 = x\sqrt{\frac{-2 \ln s}{s}}, \quad z_2 = y\sqrt{\frac{-2 \ln s}{s}}$$

и являются независимыми псевдослучайными числами, распределенными по стандартному нормальному закону $\mathcal{N}(0, 1)$.

При компьютерной реализации данного алгоритма предпочтительней использовать полярную форму, так как в этом случае приходится вычислять только одну трансцендентную функцию \ln , а не три (\ln , \sin \cos), как в стандартном варианте. Это компенсирует даже тот факт,

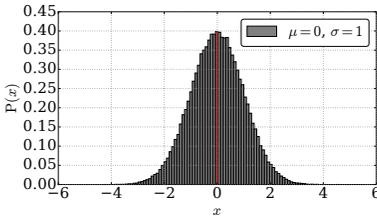


Рис. 1. Нормальное распределение

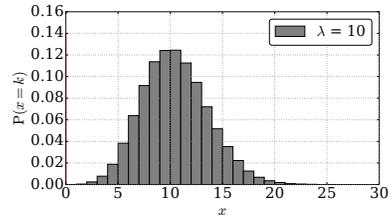


Рис. 2. Распределение Пуассона

что часть исходных равномерно распределенных числе отбрасывается — полярная версия метода все равно работает быстрее. Пример работы алгоритма изображен на рисунке 1

Для получения нормального распределения общего вида из стандартного нормального распределения используют формулу $Z = \sigma \cdot z + \mu$, где $z \sim \mathcal{N}(0, 1)$, а $Z \sim \mathcal{N}(\mu, \sigma)$.

2.2. Генерирование распределения Пуассона

Для генерирования распределения Пуассона существует большое число различных алгоритмов [4–6]. Наиболее простой был предложен Кнудом [7, 8]. Для работы алгоритманеобходимо уметь генерировать равномерные псевдослучайные числа из промежутка $[0, 1]$. Пример работы алгоритма изображен на рисунке 2

3. Определение винеровского и пуассоновского случайных процессов

Случайный кусочно-постоянный процесс $N = \{N_t, 0 \leq t \leq T\}$ с интенсивностью $\lambda > 0$ называется *процессом Пуассона* (пуассоновским, см. [9]) если выполняются следующие свойства.

1. $P\{N_0 = 0\} = 1$, иначе говоря $N_0 = 0$ почти наверное.
2. N_t — процесс с независимыми приращениями, то есть $\{\Delta N_0, \Delta N_1, \dots\}$ независимые случайные величины; $\Delta N_{t_i} = N_{t_{i+1}} - N_{t_i}$ и $0 \leq t_0 < t_1 < t_2 < \dots < t_n \leq T$; $\Delta N_{t_i} = N_{t_{i+1}} - N_{t_i}$ и $0 \leq t_0 < t_1 < t_2 < \dots < t_n \leq T$.
3. Существует число $\lambda > 0$ такое, что для любого приращения ΔN_i , $i = 0, \dots, n - 1$, $E[\Delta N_i] = \lambda \Delta t_i$.
4. Если $P(s) = P\{N_{t+s} - N_t > 2\}$, то $\lim_{s \rightarrow 0} \frac{P(s)}{s} = 0$.

Случайный процесс $W = \{W_t, 0 \leq t \leq T\}$ называется скалярным процессом Винера (винеровским, см. [9]) если выполняются следующие условия.

1. $P\{W_0 = 0\} = 1$, иначе говоря $W_0 = 0$ почти наверное.
2. W_t — процесс с независимыми приращениями, то есть $\{\Delta W_0, \Delta W_1, \dots\}$ независимые случайные величины; $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$ и $0 \leq t_0 < t_1 < t_2 < \dots < t_n \leq T$.
3. $\Delta W_i = W_{t_{i+1}} - W_{t_i} \sim \mathcal{N}(0, t_{i+1} - t_i)$ где $0 \leq t_{i+1} < t_i < T$, $i = 0, 1, \dots, n-1$.

Из определения следует, что ΔW_i — нормально распределенная случайная величина с математическим ожиданием $\mathbb{E}[\Delta W_i] = \mu = 0$ и дисперсией $\mathbb{D}[\Delta W_i] = \sigma^2 = \Delta t_i$.

Винеровский процесс является моделью броуновского движения (хаотического блуждания). Если рассмотреть процесс W_i в те моменты времени $0 = t_0 < t_1 < t_2 < \dots < t_{N-1} < t_N$ когда он испытывает случайные аддитивные изменения, то непосредственно из определения следует: $W_{t_1} = W_{t_0} + \Delta W_0$, $W_{t_2} = W_{t_1} + \Delta W_1$, \dots , $W(t_N) = W(t_{N-1}) + \Delta W_{n-1}$, где $\Delta W_i \sim \mathcal{N}(0, \Delta t_i)$, $\forall i = 0, \dots, n-1$.

Запишем W_{t_n} в виде накопленной суммы: $W_{t_n} = W_{t_0} + \sum_{i=0}^{n-1} \Delta W_i$

и учтем, что $\mathbb{E}[\Delta W_i] = 0$ и $\mathbb{D}[\Delta W_i] = \Delta t_i$ можно показать, что сумма нормально распределенных случайных чисел ΔW_i также является нормально распределенным случайным числом:

$$\mathbb{E} \sum_{i=0}^{n-1} \Delta W_i = 0, \quad \mathbb{D} \sum_{i=0}^{n-1} \Delta W_i = \sum_{i=0}^{n-1} \Delta t_i = T, \quad \sum_{i=0}^{n-1} \Delta W_i \sim \mathcal{N}(0, T).$$

Многомерный винеровский процесс $\mathbf{W}_t: \Omega \times [t_0, T] \rightarrow \mathbb{R}^m$ определяют как случайный процесс составленный из совместно независимых одномерных винеровских процессов W_t^1, \dots, W_t^m . Приращения ΔW_i^α , $\forall \alpha = 1, \dots, m$ являются совместно независимыми нормально распределенными случайными величинами. С другой стороны вектор ΔW_i^α можно представить как многомерную нормально распределенную случайную величину с вектором математических ожиданий $\mu = 1$ и диагональной матрицей ковариаций.

4. Генерирование пуассоновского и винеровского процессов в OpenModelica

4.1. Генерирование винеровского процесса

Для симулирования одномерного винеровского процесса необходимо сгенерировать N нормально распределенных случайных чисел $\varepsilon_1, \dots, \varepsilon_N$ и построить их накопленные суммы $\varepsilon_1, \varepsilon_1 + \varepsilon_2, \varepsilon_1 + \varepsilon_2 + \varepsilon_3$

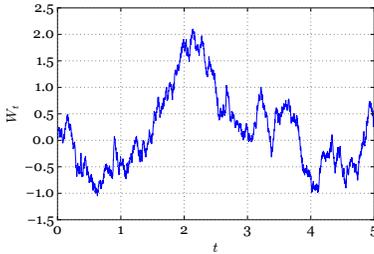


Рис. 3. Винеровский процесс

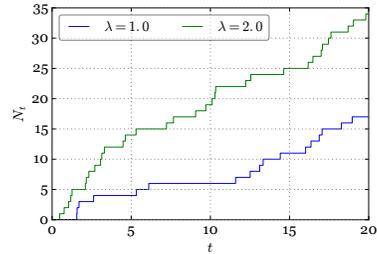


Рис. 4. Пуассоновский процесс

... В результате мы получим *выборочную траекторию* винеровского процесса $W(t)$ см. рис. 3.

В случае многомерного случайного процесса следует сгенерировать уже m последовательностей из N нормально распределенных случайных величин.

4.2. Генерирование пуассоновского процесса

Симулирование пуассоновского процесса во многом аналогично винеровскому, но теперь необходимо сгенерировать последовательность из чисел распределенных по пуассоновскому закону и затем вычислить их накопленную сумму. График процесса изображен на рис. 4. Из графика видно, что пуассоновский процесс представляет собой скачкообразное изменение числа произошедших с течением времени событий. От интенсивности λ зависит среднее количество событий за отрезок времени.

Из за такого характерного поведения пуассоновский процесс также называется скачкообразным, а стохастические дифференциальные уравнения где он участвует в качестве второго ведущего процесса получили названия уравнений со скачками [10]

4.3. Описание примера реализации в OpenModelica

Опишем реализацию генератора нормального распределения и винеровского процесса. Будем предполагать, что генератор равномерно-распределенных случайных величин уже реализован в виде функции `urand`. Для генерирования нормального воспользуемся вышеописанным преобразованием Бокса-Мюллера, а элементы последовательности, задающей винеровский процесс вычислим как кумулятивную сумму нормально-распределенных чисел.

Минимальный рабочий вариант кода приведен ниже. Ключевым моментом является использование оператора `sample(t_0, h)`, который генерирует события через h секунд, начиная с момента времени

t_0. При каждом срабатывании оператора `sample` вызывается функция `urand`, которая возвращает новое случайное число.

```

model generator
  Integer x1, x2;
  Port lg; "Порт для линейного генератора"
  Port normal; "Порт для нормально распределенных чисел"
  Port wiener; "Порт для винеровского процесса"
  Integer m = 429496729; "Значение модуля (маски)"
  Real u1, u2;
initial equation
  x1 = 114561;
  x2 = 148166;
algorithm
  when sample(0, 0.1) then
    x1 := urand(x1);
    x2 := urand(x2);
  end when;
  lg.data[1] := x1 / m;
  lg.data[2] := x2 / m;
  u1 := lcg.data[1];
  u2 := lcg.data[2];
  // нормальный генератор и винеровский процесс
  normal.data[1] := sqrt(-2 * log(u1)) * sin(6.28 * u2);
  normal.data[2] := sqrt(-2 * log(u1)) * cos(6.28 * u2);
  wiener.data[1] := wiener.data[1] + normal.data[1];
  wiener.data[2] := wiener.data[2] + normal.data[2];
end generator;

```

Отметим также применение специальной переменной типа `Port`, которая служит для соединения различных моделей между собой. В нашем примере мы создали три таких переменных: `lg`, `normal`, `wiener`. Благодаря этому, другие модели могут получить доступ к результату работы нашего генератора.

Ниже представлен минимальный код примера иллюстрирующего соединение моделей между собой. В качестве примера выбрана система из двух обыкновенных дифференциальных уравнений, описывающая осциллятор Ван-дер-Поля–Дуффинга и к ней добавлен стохастический вклад в виде процесса Винера.

$$\dot{x} = y, \quad \dot{y} = x(1.0 - x^2) - y + x \cdot W_t.$$

Отметим особо, что это уравнение не является стохастическим. Встроенные в `OpenModelica` численные методы не позволяют решать стохастические дифференциальные уравнения.

```

model ODE
  Real x, y;
  Port IN;
initial equation

```

```

    x = 2.0;
    y = 0.0;
equation
    der(x) = y ;
    der(y) = x*(1-x*x) - y + x*IN.data[1];
end ODE;
model sim
    generator gen;
    ODE eq;
equation
    connect(gen.wiener, eq.IN);
end sim;

```

5. Заключение

В статье были рассмотрены алгоритмы генерации последовательности псевдослучайных чисел, распределенных по нормальному и пуассоновскому закону, а также алгоритмы генерации элементов пуассоновского и винеровского процессов. Данные алгоритмы использовались для реализации вышеперечисленных процессов в среде OpenModelica. Был разобран минимальный работающий пример и указаны некоторые особенности его реализации.

Благодарности

Работа частично поддержана грантами РФФИ № 14-01-00628, 15-07-08795 и 16-07-00556.

Литература

1. Box G. E. P., Muller M. E. A note on the generation of random normal deviates // *The Annals of Mathematical Statistics*. — 1958. — 06. — Vol. 29, no. 2. — P. 610–611.
2. Bell J. R. Algorithm 334: Normal Random Deviates // *Commun. ACM*. — 1968. — July. — Vol. 11, no. 7. — P. 498–.
3. Knop R. Remark on algorithm 334 [g5]: Normal Random Deviates // *Commun. ACM*. — 1969. — May. — Vol. 12, no. 5. — P. 281–.
4. Devroye L. *Non-Uniform Random Variate Generation*. — New York : Springer-Verlag, 1986.
5. Ahrens J. H., Dieter U. Computer methods for sampling from gamma, beta, poisson and binomial distributions // *Computing*. — 1974. — Vol. 12, no. 3. — P. 223–246.
6. Ahrens J. H., Dieter U. Computer generation of poisson deviates from modified normal distributions // *ACM Trans. Math. Softw.* — 1982. — June. — Vol. 8, no. 2. — P. 163–179.

7. Кнут, Дональд Эрвин. Искусство программирования. — 3 изд. — Москва : Вильямс, 2004. — Т. 2.
8. Knuth D. E. The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms. — Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 1997. — Vol. 2.
9. Kloeden P. E., Platen E. Numerical Solution of Stochastic Differential Equations. — 2 edition. — Berlin Heidelberg New York : Springer, 1995.
10. Platen E., Bruti-Liberati N. Numerical Solution of Stochastic Differential Equations with Jumps in Finance. — Heidelberg Dordrecht London New York : Springer, 2010.

UDC 004.94+519.62

Implementation of the Wiener stochastic process in OpenModelica

M. N. Gevorkyan*, A. V. Demidova*, A. V. Korolkova*,
D. S. Kulyabov*[†], L. A. Sevastianov*[‡]

* *Department of Applied Probability and Informatics
RUDN University*

Miklukho-Maklaya str. 6, Moscow, 117198, Russia

[†] *Laboratory of Information Technologies
Joint Institute for Nuclear Research*

Joliot-Curie 6, Dubna, Moscow region, 141980, Russia

[‡] *Bogoliubov Laboratory of Theoretical Physics
Joint Institute for Nuclear Research*

Joliot-Curie 6, Dubna, Moscow region, 141980, Russia

In an open implementation of component-oriented modeling language OpenModelica there are no tools to simulate stochastic processes. The article considers the problem of adding generators random distributions that are not uniform, in particular the normal and Poisson.. The normal distribution is then used for the simulation of a Wiener random process, which plays a key role in stochastic differential equations.

Keywords: pseudo-random numbers, stochastic processes, Wiener process, Modelica.

УДК 004.94+519.62

Особенности расширения системы OpenModelica

М. Н. Геворкян*, А. В. Демидова*, А. В. Королькова*,
Д. С. Кулябов*[†], Л. А. Севастьянов*[‡]

* Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198

† Лаборатория информационных технологий,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980

‡ Лаборатория теоретической физики,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980

Аннотация. Система OpenModelica является открытой реализацией языка компонентно-ориентированного моделирования Modelica и основывается на множестве открытых библиотек и компиляторе C/C++ из GNU Compiler Collection. Целью данной заметки является краткое описание основных компонентов OpenModelica, используемых в этой системе библиотек, а также пошаговое руководство по использованию внешних функций, написанных на языке C, в коде OpenModelica. Используется сборка OpenModelica 1.11.0-dev-15 для дистрибутива Ubuntu Linux версии 16.04. В работе кратко описаны основные шаги, которые необходимо совершить пользователю для успешного вызова C-функций непосредственно из программ OpenModelica. Указаны некоторые особенности, которые отсутствуют в официальной документации.

Ключевые слова: Modelica, OpenModelica, моделирование.

1. Введение

Данная статья посвящена обзору возможностей расширения системы OpenModelica [1] [2] с помощью вызова внешних функций, написанных на языке C. Кроме того описываются основные компоненты данной системы и внешние библиотеки, которые используются для численного решения уравнений, которые являются основой практически любой модели. Выбор той или иной процедуры для численного решения

2. Основные компоненты OpenModelica

Первая версия языка Modelica была опубликована в 1997 году. В 2000 году была создана некоммерческая ассоциация, целью которой является развитие стандарта языка и стандартной библиотеки. Начиная с 2002 года проводится регулярная конференция, по результатам которой публикуются материалы, часть из которых доступна на официальном сайте. Ассоциация Modelica разрабатывает лишь стандарт, поэтому существует несколько различных реализаций этого стандарта

2.1. Реализации языка Modelica

Кроме свободной реализации OpenModelica существует большое количество коммерческих реализаций языка Modelica, включенных в пакеты программ для инженерных и математических расчетов: LMS Imagine.Lab Amesim [3], Dymola [4], Wolfram System Modeler [5] и MampleSim [6]. Кроме OpenModelica существуют другие свободные реализации: JModelica.org [7] и Modelicas в составе SciLab [8]. Полный обновляемый список программного обеспечения доступен на официальном сайте ассоциации Modelica [2].

2.2. Компоненты OpenModelica

В состав OpenModelica входит около десяти компонент. Кратко опишем основные из них (полный список доступен на официальном сайте [1] в разделе Tools and Apps).

- Advanced Interactive OpenModelica Compiler (ОМС) — центральный компонент всего пакета программ. Предназначен для компиляции modelica-кода в код на языке Си.
- Интерактивная оболочка OMSshell — выполняет роль интерактивной консоли для выполнения команд Modelica.
- OMNotebook — интерактивная среда для выполнения фрагментов кода Modelica с возможностью отображения графиков. Создан по аналогии с интерактивными оболочками Maple, Mathematica, iPython и других.
- OMEdit — редактор для исходного кода, а также для визуального программирования путем соединения блоков между собой. Непосредственно из OMEdit можно запустить компиляцию созданных моделей и просмотреть результаты моделирования в виде графиков, зависящих от времени.

Основным средством для написания программ на языке Modelica может служить OMEdit, который по сути является интегрированной средой разработки и поддерживает отладку, подсветку синтаксиса, автоматическое форматирование и т.д. Интерактивная оболочка OMNotebook удобна для экспериментирования с частями программы, в случае если четкий алгоритм работы программы еще не ясен, так как OMNotebook позволяет интерактивно выполнять фрагменты кода, сразу же получая результаты их работы.

3. Встроенные численные методы

Основным содержимым классов в программе, написанной на языке Modelica, являются системы линейных, нелинейных и дифференциальных уравнений. Специализированный процессор анализирует заданную систему уравнений и определяет порядок вычисления переменных и используемый численный метод. Выбор численного метода не

доступен пользователю в явном виде и скрыт в настройках симуляции. Опишем основные из этих методов и дадим ссылки на используемые открытые библиотеки.

Для численного решения обыкновенных дифференциальных уравнений в OpenModelica используются следующие численные методы.

- Метод Эйлера (euler) [9, 10, §I.7].
- Метод Рунге–Кутты 4-го порядка (rungekutta) [9, 10, §II.1].
- Метод Адамса–Мультона 1-5 порядков (dassl) [9, 10, §III.5].
- Метод Радо ПА (radau1,3,5) и метод Лобатто ПА (lobatto2,4,6) [9, 10, §II.7].

Метод Адамса–Мультона используется по умолчанию и это единственный метод с управлением шагом.

Для решения систем линейных уравнений используются процедуры из различных библиотек, таких как lapack [11], lis [12], umfpack [13], totalpivot. Для нелинейных систем: hybrid, kinsol [14], newton, homotopy, mixed.

Результат вычислений по умолчанию сохраняется во временной директории в виде файла формата `mat`. Данный формат бинарный, является родным для пакета математических вычислений MATLAB и используется для хранения матричных данных. В окне настройки симуляции можно дополнительно выбрать `plt` и `csv` форматы. Формат `plt` разработан компанией HP для печати на плоттерах и чаще всего используется в программе AutoCAD. Формат `csv` наиболее простой из вышеперечисленных. Он является текстовым и представляет собой таблицу данных, разделенных запятыми или иным разделителем.

Знание используемых численных методов важно, так как каждый из этих методов имеет свою область применения и может работать некорректно в случае специфической системы уравнений. В случае прямолинейного применения средства моделирования Modelica могут давать некорректные результаты. Так например, в случае необходимости численного интегрирования систем уравнений на длительном промежутке времени может возникнуть необходимость в использовании геометрических численных методов [15, 16] для предотвращения накопления ошибки. Из изложенного выше следует, что OpenModelica такие методы не поддерживает.

4. Подключение внешних C функций к OpenModelica

Перейдем к рассмотрению процесса подключения внешних функций к программе на OpenModelica. Соответствующий раздел в официальной документации при описании данной процедуры упускает несколько существенных деталей и следование его указаниям приводит к ошибке. Все описанные нами действия проводились на компьютере под управлением операционной системы Ubuntu Linux версии 16.04 LTS и OpenModelica версии 1.11.0-dev-15.

При компиляции кода OpenModelica программа транслируется в С-код, который затем обрабатывается уже С-компилятором. Поэтому поддержка С-функций реализована в OpenModelica на уровне архитектуры. Кроме языка С OpenModelica поддерживает также вызов Fortran (только F77) и Python функций. Однако оба этих языка поддерживаются косвенным путем, а именно с помощью обертывания их в соответствующие С-функции.

Использование внешних С-функций может понадобиться по разным причинам: повышение быстродействия ресурсоёмких компонент программы, необходимость использования полноценного императивного языка программирования, использование уже существующего С-кода т.д.

Приведем простейший пример вызова С функций из программы написанной на языке Modelica. Создадим два исходных файла ExternalFunc1.c и ExternalFunc2.c. В этих файлах будут содержаться простейшие функции, которые мы хотим использовать в нашей программе на языке Modelica.

```
// Файл ExternalFunc1.c
double ExternalFunc1_ext(double x)
{
    return x+2.0*x*x;
}
// Файл ExternalFunc2.c
double ExternalFunc2(double x)
{
    return (x-1.0)*(x+2.0);;
}
```

В директории, где находится исходный код программы на языке Modelica необходимо создать каталог Resources а в нем каталог Library, в котором будут находиться оба наших файла ExternalFunc1.c и ExternalFunc2.c. После этого следует создать объектные файлы и поместить их в архив, который будет служить подключаемой библиотекой. Для этого следует выполнить ряд команд.

```
gcc -c -o ExternalFunc1.o ExternalFunc1.c
gcc -c -o ExternalFunc2.o ExternalFunc2.c
ar rcs libExternalFunc1.a ExternalFunc1.o
ar rcs libExternalFunc2.a ExternalFunc2.o
```

Для создания объектных файлов мы использовали утилиту компиляции gcc с опцией -c и архиватор ar для помещения созданных объектных файлов в архив. В результате мы получим два файла libExternalFunc1.a и libExternalFunc2.a. Есть также возможность поместить все необходимые нам объектные файлы в один архив.

Для вызова внешних функций следует использовать ключевое слово external. Название функции-обертки на языке Modelica может как

совпадать, так и отличаться от названия внешней функции. Во втором случае необходимо явно указать какую внешнюю функции следует обернуть.

```

model ExternalLibraries
// Название функции не совпадает с названием функции
// на языке C
function ExternalFunc1
  input Real x;
  output Real y;

// Явно указываем название C-функции
external y=ExternalFunc1_ext(x)
  annotation(Library="ExternalFunc1");
end ExternalFunc1;

function ExternalFunc2
input Real x;
output Real y;
// Названия функций совпадают, поэтому явное указание
// не требуется
external "C" annotation(Library="ExternalFunc2");
end ExternalFunc2;

Real x(start=1.0, fixed=true), y(start=2.0, fixed=true);
equation
  der(x)=-ExternalFunc1(x);
  der(y)=-ExternalFunc2(y);
end ExternalLibraries;

```

Заметим, что в аннотации название подключаемой библиотеки указано как `ExternalFunc1`, в то время как сам файл называется `libExternalFunc1.a`. Это не является опечаткой и приставку `lib` необходимо добавлять ко всем библиотечным файлам.

Из примера видно, что тип `Real` языка `Modelica` соответствует типу `double` языка `C`. Кроме того тип `Integer` и `Boolean` соответствуют типу `int`. Массивы типа `Real` и `Integer` также переводятся в массивы типа `double` и `int`.

Стабильно работает вызов функций с аргументами типа `int` и `double`, а также массивами этих типов. Попытка же уточнить используемый тип, например `long long int` или использовать беззнаковый тип, например `unsigned int` приводит к ошибке.

5. Заключение

В статье были описаны используемые в `OpenModelica` численные методы и библиотеки. Эти сведения позволят пользователю, планирующему использовать `OpenModelica`, заранее решить сумеет ли

OpenModelica справится с корректным решением используемых для моделирования уравнений. вызова внешних функций из программы, написанной на языке OpenModelica. Также подробно освещена процедура подключения внешних C-функций к программе на OpenModelica. Описаны некоторые пункты, опущенные в официальной документации.

Благодарности

Работа частично поддержана грантами РФФИ № 14-01-00628, 15-07-08795 и 16-07-00556.

Литература

1. OpenModelica official site. — URL: <https://www.openmodelica.org/> (online; accessed: 19.06.2016).
2. Modelica and the Modelica Association official site. — URL: <https://www.modelica.org/> (online; accessed: 19.06.2016).
3. LMS Imagine.Lab Amesim. — URL: http://www.plm.automation.siemens.com/en_us/products/lms/imagine-lab/amesim/index.shtml (online; accessed: 19.06.2016).
4. Multi-Engineering Modeling and Simulation - Dymola - CATIA. — URL: <http://www.3ds.com/products-services/catia/products/dymola> (online; accessed: 19.06.2016).
5. Wolfram SystemModeler. — URL: <http://www.wolfram.com/system-modeler/index.ru.html> (online; accessed: 19.06.2016).
6. MapleSim - High Performance Physical Modeling and Simulation - Technical Computing Software. — URL: <http://www.maplesoft.com/products/maplesim/index.aspx> (online; accessed: 19.06.2016).
7. Jmodelica.org. — URL: <http://www.jmodelica.org/> (online; accessed: 19.06.2016).
8. SciLab. — URL: <http://www.scilab.org/> (online; accessed: 19.06.2016).
9. Hairer E., Nørsett S. P., G.Wanner. Solving Ordinary Differential Equations I. — 2 edition. — Berlin : Springer, 2008.
10. Хайрер Э. and Нёрсетт С. and Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи / Под ред. С. С. Филиппова. — 1 изд. — Москва : Мир, 1990.
11. LAPACK — Linear Algebra PACKage. — URL: <http://www.netlib.org/lapack/> (online; accessed: 19.06.2016).
12. Nishida A., Fujii A., Oyanagi Y. Lis: Library of Iterative Solvers for Linear Systems. — URL: http://www.phy.duke.edu/~rgb/General/rand_rate.php (online; accessed: 19.06.2016).

13. SuiteSparse : a suite of sparse matrix software. — URL: <http://faculty.cse.tamu.edu/davis/suitesparse.html> (online; accessed: 19.06.2016).
14. Collier A. M., Hindmarsh A. C., Serban R., dward C. S. W. User Documentation for KINSOL v2.8.2. — 2015. — URL: http://computation.llnl.gov/sites/default/files/public/kin_guide.pdf.
15. Hairer E., Lubich C., Wanner G. Geometric numerical integration: structure-preserving algorithms for ordinary differential equations. Springer series in computational mathematics. — Springer, 2006.
16. Sanz-Serna J., Calvo M. Numerical Hamiltonian Problems. — 1 edition. — London : Chapman and Hall, 1994.

UDC 004.94+519.62

About extensions programming for OpenModelica

M. N. Gevorkyan*, A. V. Demidova*, A. V. Korolkova*,
D. S. Kulyabov*[†], L. A. Sevastianov*[‡]

* *Department of Applied Probability and Informatics
RUDN University*

Miklukho-Maklaya str. 6, Moscow, 117198, Russia

[†] *Laboratory of Information Technologies*

Joint Institute for Nuclear Research

Joliot-Curie 6, Dubna, Moscow region, 141980, Russia

[‡] *Bogoliubov Laboratory of Theoretical Physics*

Joint Institute for Nuclear Research

Joliot-Curie 6, Dubna, Moscow region, 141980, Russia

OpenModelica is an open implementation of Modelica — the component-oriented modeling language. It is based on many open source libraries and C/C++ compiler from the GNU Compiler Collection. The goal of this short paper is a brief description of the main components of the OpenModelica, libraries, which are used in it , as well as step-by-step guidance on using external functions written in C We use OpenModelica build 1.11.0-dev-15 for Ubuntu 16.04 LTS. This short paper briefly describes the main steps that a user need to implement for successful call of C-functions directly from OpenModelica code. In addition, some features are specified which do not exist in official documentation.

Keywords: Modelica, OpenModelica, modelling.

УДК 519.24

Моделирование плотности распределения цен на бирже в виде произведения процессов Вейбулла

И. В. Глухов*, Ю. Н. Орлов*†

* Кафедра высшей математики,
Московский физико-технический институт (государственный
университет),
Институтский пер., д.9, Долгопрудный, Московская обл., Россия, 141700

† Сектор кинетических уравнений,
Институт прикладной математики имени Келдыша,
Миусская пл., 4, Москва, Россия, 125047

Аннотация. Моделируется плотность распределения цен некоторых финансовых инструментов ММВБ. Показано, что построенные модели адекватно моделируют распределение цен ценных бумаг. Модельные плотности распределения построены в виде нормализованного произведения двух плотностей распределения процессов Вейбулла. Посчитаны погрешности модели в зависимости от количества интервалов выборки и сравнены с статистической точностью гистограммы.

Ключевые слова: плотность распределения, распределение Вейбулла, статистическая гистограмма, статистическая точность гистограммы.

1. Введение

В работе рассмотрена модель распределения с плотностью:

$$f_{model}(x) = \frac{f_1(x)f_2(x)}{\int_{-\infty}^{+\infty} f_1(t)f_2(t)dt}, \quad (1)$$

Где $f_1(x)$ и $f_2(x)$ - плотности табличных распределений. В начале в работе плотность распределения цен была смоделировано в виде произведения плотностей нормальных распределений. В этом случае модельное распределение тоже является нормальным и данная задача уже была решена в прошлом аналитически.

2. Основная часть

В этой работе работе была рассмотрена модель для распределения отклонения цен на некоторый финансовый инструмент в течение "пятиминутки". Для выборки была построена статистическая гистограмма $f_{practice}(x)$. В качестве теоретической модели данной гистограммы была рассмотрена плотность распределения (1), где в качестве $f_1(x)$ и $f_2(x)$ были взяты плотности распределений Вейбулла. Плотность

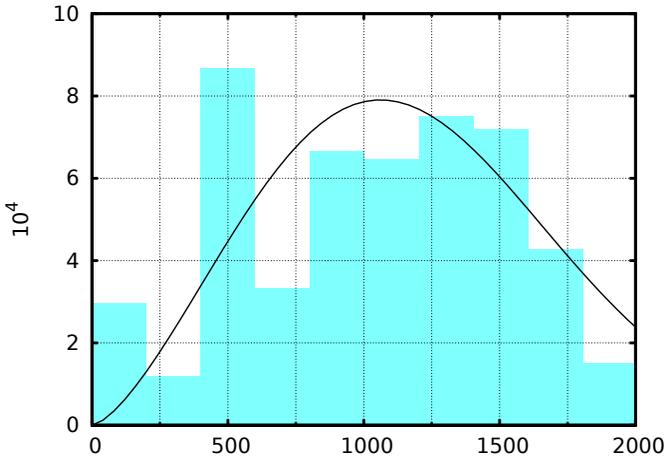


Рис. 1. Нормализованная гистограмма, теоретическая плотность распределения

распределения Вейбулла (двухпараметрическое распределение с параметрами λ и k) задается формулой:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & \text{если } x \geq 0; \\ 0, & \text{если } x < 0. \end{cases}$$

С помощью программы были посчитаны оптимальные параметры $\lambda_1, k_1, \lambda_2, k_2$ для $f_1(x)$ и $f_2(x)$, такие что норма $\|f_{model} - f_{practice}\|_{RL^1}$ является минимальной.

Для данной модели в работе была посчитана зависимость погрешности модели от количества интервалов выборки в статистической гистограмме. По этим подсчетам было выявлено оптимальное количество интервалов $n_{practice}$ выборки в статистической гистограмме и сравнено с n , которое статистической точностью диаграммы n , найденным из формулы статистической точности гистограммы:

$$\frac{t_{1-\epsilon}(N-1)}{\epsilon} = \sqrt{\frac{N}{n}},$$

где $t_{1-\epsilon}(N-1)$ - квантиль распределения Стьюдента.

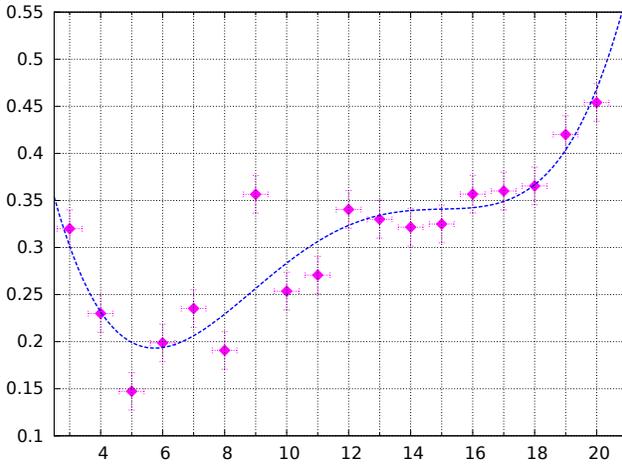


Рис. 2. Зависимость погрешности модели от количества интервалов выборки гистограммы

3. Заключение

Таким образом, в результате исследования решены следующие задачи:

- установлена работоспособность моделирования распределения ценных бумаг произведением Вейбулловских процессов.
- посчитано оптимальное количество интервалов выборки практическим и теоретическим способами.
- построенные аддитивные адекватные модели применимы для прогноза стоимости ценных бумаг.

Литература

1. Орлов Ю. Нестационарные временные ряды [Текст] : методы прогнозирования с примерами анализа финансовых и сырьевых рынков. — URSS, 2011. — ISBN: 9785397015417.
2. Айвазян С., Енюков И., Мешалкин Л. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное издание — Москва: Финансы и статистика, 1983. — 471 стр.

UDC 519.24

Modeling of stock prices in the density distribution of a product of processes Weibull

I. V. Glukhov*, Y. N. Orlov*[†]

** Department of High Mathematics
Moscow Institute of Physics and Technology (State University)
Institutskiy per. 9, Dolgoprudny, Moscow Region, 141700, Russia*

*[†] Sector of Kinetic Equations
Keldysh Institute of Applied Mathematics
Miusskaya sq. 4, Moscow, 125047, Russia*

Density distribution of the price of certain financial instruments MICEX was modeled. It is shown that the model adequately model the distribution of the price of securities. Model of density distribution was constructed in the form of a normalized multiplication of two Weibull distribution densities. Model's errors were calculated depending on the number of sample intervals and compared to the statistical accuracy of the histogram.

Keywords: density distribution , Weibull distribution , statistical histogram , statistical accuracy of the histogram.

УДК 519.218.31

Математическая модель системы облачных вычислений

А. В. Горбунова*, А. А. Краденых*, И. С. Зарядов*†

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

† *Институт проблем информатики ФИЦ ИУ РАН,
ул. Вавилова 44 кор.2, ИПИ ФИЦ ИУ РАН, Москва, Россия, 119333*

Аннотация. Исследуется система облачных вычислений с помощью системы массового обслуживания с расщеплением поступающих запросов на части и синхронизацией обслуживания. Запросы поступают в систему как единое целое и только при переходе на обслуживание разбиваются на части, каждая часть занимает свой прибор, и при этом приборы, обслуживающие части одной и той же заявки, считаются занятыми до тех пор, пока не обслужится весь запрос целиком (в зависимости от политики обслуживания). Рассматриваются две политики обслуживания: запрос считается обслуженным тогда, когда обслужились все его части (время отклика — максимум времён обслуживания всех частей данной заявки); либо запрос считается обслуженным тогда, когда обслужится хотя бы одна его часть (как только одна из частей обслужилась, все остальные части также покидают свои приборы и новый запрос переходит на обслуживание). В этом случае время обслуживания одного запроса есть минимум времён обслуживания всех его частей. Получены характеристики функционирования системы для обоих представленных вариантов.

Ключевые слова: система облачных вычислений, система массового обслуживания, время отклика, вероятностные характеристики.

1. Введение

Облачные вычисления (англ. cloud computing) представляют собой информационно-технологическую концепцию, подразумевающую обеспечение удаленного сетевого доступа к различным ресурсам (программам, данным, вычислительным мощностям) посредством интернет-сервисов. Среди основных достоинств облачных вычислений можно перечислить доступность, гибкость, легкость в управлении, сравнительно низкую стоимость внедрения [1].

Одна из важнейших характеристик при функционировании систем облачных вычислений — это время отклика [?, 1, 5, 8]. То есть, время, за которое заявка, разбитая на части, считается обслуженной полностью. Таким образом, в зависимости от типа задачи, время отклика определяется либо как максимум времён обслуживания частей запроса, либо как минимум.

В зависимости от типа решаемых задач, поступающая на обслуживание заявка (задача), разбивается:

- на k одинаковых подзапросов, и тогда запрос считается обслуженным целиком, как только обслужится хотя бы одна его часть (подзапрос). В этом случае, время отклика определяется как минимум времен обслуживания всех подзапросов.
- на k подзапросов, каждый из которых должен быть обработан (не обязательно последовательно). Тогда запрос (задача) считается обработанным полностью только тогда, когда обработаны все подзапросы. В этом случае время отклика есть максимум времен обслуживания каждого из подзапросов одного и того же запроса [2, 2–5, 8–10].

Математические модели моделирования систем облачных вычислений с помощью теории массового обслуживания можно разбить на следующие виды:

1. Представление системы облачных вычислений в виде системы массового обслуживания (СМО), состоящей из нескольких подсистем, в каждой из которых есть свой обслуживающий прибор и свой накопитель. Поступающий запрос разбивается на k частей, каждая часть становится в свою подсистему (системы типа Fork–Join или Split–Merge) [2, 4, 5, 8–11].
2. Система облачных вычислений представляется как СМО, состоящая из бесконечного числа обслуживающих приборов. Поступает поток запросов, запросы разбиваются на части, и каждая часть занимает свой прибор [6, 7].

Предлагаемая в данной работе математическая модель является "синтезом" модели Split–Merge (синхронизация подзапросов) и модели с бесконечным числом приборов. Наша математическая модель представляет собой СМО, состоящую из накопителя и нескольких обслуживающих приборов (виртуальных машин). В накопитель поступает поток требований, каждое требование в накопителе занимает одно место, как единое целое, и разбивается на подзапросы только при переходе на приборы. Каждый подзапрос занимает свой прибор и обслуживается на нем согласно заданному распределению времени обслуживания. Приборы, обслуживающие части одного и того же запроса, считаются занятыми до тех пор, пока не обслужились все части этого подзапроса (если иное не обговорено в политике обслуживания). Следовательно, новый запрос может перейти на обслуживание только тогда, когда либо есть достаточное число свободных приборов для его обслуживания, либо обслужились все подзапросы предыдущего запроса (реализация синхронизации механизма Split–Merge). Если число приборов стремится к бесконечности, то полученные результаты будут аналогичны результатам работ [6, 7].

2. Описание системы

Рассматривается система массового обслуживания (СМО), состоящая из накопителя ёмкости $r < \infty$ и k обслуживающих приборов, времена обслуживания на которых независимы и имеют экспоненциальные распределения с интенсивностями μ_i ($i = \overline{1, k}$). В рассматриваемую систему массового обслуживания поступает пуассоновский поток запросов интенсивности λ . Каждый запрос состоит из k частей и занимает одно место в накопителе. При переходе на обслуживание запрос расщепляется и каждая его часть обслуживается на своём приборе.

Две политики обслуживания — запрос считается обслуженным на приборах, если

- обработана хотя бы одна его часть (подзапрос) (минимум времён обслуживания всех частей запроса).
- обработаны все подзапросы (максимум времён обслуживания каждого из подзапросов одного и того же запроса) [8–10].

Для построения математической модели будем использовать многомерный марковский процесс $\nu(t) = (\xi(t), \vec{\varphi}(t))$, где $\xi(t)$ — число запросов в системе в момент времени t , а $\vec{\varphi}(t) = (\varphi_1(t), \dots, \varphi_k(t))$, — вектор, описывающий число частей запроса (0 или 1) на каждом из k приборов в момент времени t . В предположении о существовании стационарного режима обозначим через $p_{l;(\overline{i_1, \dots, i_k})} = P\{\xi(t) = l, \vec{\varphi}(t) = (\overline{i_1, \dots, i_k})\}$ ($l = \overline{1, r+1}$, $i_j = \overline{0, 1}$, $j = \overline{1, k}$) вероятность того, что в системе ровно l запросов и хотя бы один прибор занят. Кроме того, через $p_0 = p_{0;(\overline{0, \dots, 0})}$ обозначим вероятность того, что система пуста.

3. Время отклика как максимум

Определим время обслуживания (время отклика) одного запроса η как максимум времён обслуживания всех частей этого запроса, т.е. $\eta = \max(\eta_1, \eta_2, \dots, \eta_k)$. Тогда функция распределения времени обслуживания имеет вид:

$$P\{\max(\eta_1, \eta_2, \dots, \eta_k) < x\} = \prod_{i=1}^k (1 - e^{-\mu_i x}).$$

Рассмотрим частный случай, когда в системе $k = 2$ приборов.

Введём вспомогательные обозначения: $p_l = p_{l;(\overline{1, 1})}$ — вероятность того, что в системе ровно l ($l = \overline{1, r+1}$) запросов и все обслуживающие приборы заняты частями одного запроса, $\tilde{p}_l = p_{l;(\overline{1, 0})} + p_{l;(\overline{0, 1})}$ — вероятность того, что в системе ровно l ($l = \overline{1, r+1}$) запросов и один прибор свободен.

Система уравнений равновесия (СУР) для рассматриваемого случая:

$$\begin{aligned}\lambda p_0 &= \mu_1 p_{1;(1,0)} + \mu_2 p_{1,(0,1)}, \\ (\lambda + \mu_1 + \mu_2) p_i &= \lambda p_{i-1} + \mu_1 p_{i+1;(1,0)} + \mu_2 p_{i+1,(0,1)}, \quad i = \overline{1, r}, \\ (\mu_1 + \mu_2) p_{r+1} &= \lambda p_r, \\ (\lambda + \mu_1) p_{1;(1,0)} &= \mu_2 p_1, \quad (\lambda + \mu_2) p_{1;(0,1)} = \mu_1 p_1, \\ (\lambda + \mu_1) p_{l;(1,0)} &= \lambda p_{l-1;(1,0)} + \mu_2 p_l, \quad l = \overline{2, r}, \\ (\lambda + \mu_2) p_{l;(0,1)} &= \lambda p_{l-1;(0,1)} + \mu_1 p_l, \quad l = \overline{2, r}, \\ \mu_2 p_{r+1;(0,1)} &= \lambda p_{r;(0,1)} + \mu_1 p_{r+1}, \quad \mu_1 p_{r+1;(1,0)} = \lambda p_{r;(1,0)} + \mu_2 p_{r+1}.\end{aligned}$$

Условие нормировки:

$$\sum_{l=0}^{r+1} (p_{l;(1,1)} + p_{l;(1,0)} + p_{l;(0,1)}) = 1.$$

Если предположить, что $\mu_1 = \mu_2 = \mu$, то представленную выше СУР можно записать в виде:

$$\begin{aligned}\lambda p_0 &= \mu \tilde{p}_1, \quad 2\mu p_{r+1} = \lambda p_r, \\ (\lambda + 2\mu) p_i &= \lambda p_{i-1} + \mu \tilde{p}_{i+1}, \quad i = \overline{1, r}, \\ (\lambda + \mu) \tilde{p}_1 &= 2\mu p_1, \quad (\lambda + \mu) \tilde{p}_l = \lambda \tilde{p}_{l-1} + 2\mu p_l, \quad l = \overline{2, r}, \\ \mu \tilde{p}_{r+1} &= \lambda \tilde{p}_r + 2\mu p_{r+1}.\end{aligned}$$

Рекуррентные формулы для расчета вероятностей p_i ($i = \overline{0, r+1}$) и \tilde{p}_i ($i = \overline{1, r+1}$):

$$\begin{aligned}\tilde{p}_1 &= \rho p_0, \quad p_1 = 0,5\rho(\rho = 1)p_0, \\ \tilde{p}_i &= (\rho + 2)p_{i-1} - \rho p_{i-2}, \quad i = \overline{2, r}, \\ p_i &= 0,5(\rho + 1)\tilde{p}_i - 0,5\rho\tilde{p}_{i-1}, \quad i = \overline{2, r}, \\ p_{r+1} &= 0,5\rho p_r, \quad \tilde{p}_{r+1} = 2p_{r+1} + \rho\tilde{p}_r,\end{aligned}$$

где $\rho = \lambda/\mu$.

Вероятность простоя одного из двух приборов равна

$$\sum_{l=1}^{r+1} \tilde{p}_l = \frac{\lambda}{\mu} (1 - p_{r+1}^*),$$

где p_{r+1}^* — вероятность переполнения системы ($p_{r+1}^* = p_{r+1} + \tilde{p}_{r+1}$).

Для системы с бесконечным накопителем эта же вероятность равна:

$$\sum_{l=1}^{r+1} \tilde{p}_l = \frac{\lambda}{\mu}.$$

4. Время отклика как минимум

Определим время отклика η (время обслуживания запроса) как минимум времён обслуживания подзапросов на приборах, т.е. $\eta = \min(\eta_1, \eta_2, \dots, \eta_k)$. Здесь η_j — это время обслуживания j -й части запроса на j -м приборе. В предположении об экспоненциальном распределении времён обслуживания на приборах с интенсивностями μ_j , $j = \overline{1, k}$, функция распределения времени отклика имеет вид:

$$P\{\min(\eta_1, \eta_2, \dots, \eta_k) < x\} = 1 - e^{-(\mu_1 + \dots + \mu_k)x}.$$

В отличие от предыдущего случая (время отклика — максимум) в системе всегда заняты все приборы (если есть хотя бы один запрос), то логично рассматривать стационарное распределение p_i вероятностей того, что в системе находится i заявок, $i = \overline{0, r+1}$, которое имеет вид:

$$p_i = \rho^i (1 - \rho), \quad i = \overline{0, r+1}, \quad \rho = \frac{\lambda}{\sum_{j=1}^k \mu_j}.$$

Таким образом, рассматриваемая система эквивалентна системе $M|M|1|r$.

Если число приборов в системе равняется αk (α — целое неотрицательное число), то есть кратно количеству частей запроса, то стационарное распределение числа запросов в системе имеет вид:

$$p_i = \frac{\rho^i}{i!} p_0, \quad i = \overline{0, \alpha},$$

$$p_i = \frac{\rho^{\alpha+1}}{\alpha^{i-\alpha} \alpha!} p_0, \quad i = \overline{\alpha+1, \alpha+r},$$

$$p_0 = \left(\sum_{i=0}^{\alpha-1} \frac{\rho^i}{i!} + \frac{\frac{\rho^\alpha}{\alpha!} \left(\left(\frac{\rho}{\alpha} \right)^{r+1} - 1 \right)}{\frac{\rho}{\alpha} - 1} \right)^{-1}.$$

5. Заключение

Рассмотрена математическая модель системы облачных вычислений, для построения которой использовалась система массового обслуживания с расщеплением поступающих заявок для дальнейшего параллельной обработки на приборах. Представлено два варианта обслуживания запросов — запрос считается обслуженным тогда, когда обслужились все его части, и, второй вариант, запрос считается обслуженным, как только обслужилась одна из его частей. Для первого

варианта предложен рекуррентный алгоритм расчёта стационарного распределения числа заявок в системе. Для второго показано, что решение аналогично уже известному для системы $M|M|1|g$

Благодарности

Работа частично поддержана грантами РФФИ № 14-07-00090, № 15-07-03007, № 15-07-03406 и № 16-07-00766.

Литература

1. *R. Buyya, J. Broberg, and A. M. Goscinski* "Introduction to cloud computing," Cloud Computing: Principles and Paradigms, John Wiley & Sons Inc., 2011, pp. 3–42.
2. *Nelson R., Tantawi A. N.* Approximate analysis of fork–join synchronization in parallel queues. // IEEE Transactions on Computers. — 1988. — Vol. 37. — P. 739–743.
3. *Baccelli F., Makowski A. M., Shwartz A.* The fork–join queue and related systems with synchronization constraints: Stochastic ordering and computable bounds. // Advances in Applied Probability. — 1989. — Vol. 21, no. 3. — P. 629–660.
4. *Ko S. S., Serfozo R. F.* Response Times in M/M/s Fork-Join Networks. // Adv. in App. Prob. — 2004. — Vol. 36, no. 3. — P. 854–871.
5. *Xiong K., Perros H.* Service performance and analysis in cloud computing. // IEEE World Conference on Services. — 2009. — Los Angeles. — P. 693–700.
6. *И. А. Ивановская, С. П. Моисеева* Исследование математической модели параллельного обслуживания заявок смешанного типа, Известия Томского политехнического университета. Управление, вычислительная техника и информатика, т. 317, № 5, 2010, с. 32–34
7. *Ивановская И. А., Моисеева С. П.* Исследование модели параллельного обслуживания сдвоенных заявок в нестационарном режиме. // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. — 2010. — № 3 (12). — С. 21–28.
8. *Basharin G., Gaidamaka Yu., Samouylov K.* Mathematical theory of teletraffic and its application to the analysis of multiservice communication of next generation networks. // Automatic Control and Computer Sciences/ — 2013. — Vol. 47, issue 2. — P. 62–69.
9. *Mokrov E., Samouylov K.* Modeling of cloud computing as a queuing system with batch arrivals // T-Comm. — 2013, — № 11. — P. 139–141.
10. *Thomasian A.* Analysis of fork-join and related queueing systems. // ACM Comput. Surv. (CSUR). — 2014. — Vol. 47, no. 2, article 17. — P. 17.1–17.71

11. Горбунова А. В., Зарядов И. С., Матюшенко С. И., Самуйлов К. Е., Шоргин С. Я. Аппроксимация времени отклика системы облачных вычислений. // Информатика и ее применение. — Т. 9, выпуск 3. — 2015. — С. 32–38.

UDC 519.218.31

The mathematical model of a cloud computing systeme

A. V. Gorbunova*, A. A. Kradenyh*, I. S. Zaryadov*[†]

* *Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

[†] *Institute of Informatics Problems, FRC CSC RAS
PI FRC CSC RAS, 44-2 Vavilova Str., Moscow 119333, Russia*

The mathematical model of cloud computing system based on the queuing system with the splitting of the incoming queries and synchronization of services is considered. The queueing system consists of a single buffer and several servers, service times are independent and exponentially distributed. The incoming query enters the system as a whole and only before service is divided into subqueries, each subquery is served by its device. The servers with parts of the same query are considered to be employed as long as the query is not serviced as a whole (according to service policy). Two variants of service policy are presented: the query is handled only when the last of it is out and a new query may be served only when there are enough free servers (the response time is the maximum of service times of all parts of this query) or the query is handled when at least one part of it is serviced (as only one subquery is served, all others leave their servers and a new query from the buffer goes to the service). In this case, the response time is the minimum of service times of its parts. The stationary probability-time characteristics of the system are presented for both mentioned options.

Keywords: cloud computing system, queuing system, control, computation, communication, probability-time characteristics.

UDC 519.25

On noising of data to refine the output of moving separation of mixtures

A. K. Gorshenin*, V. Yu. Korolev†*

** Institute of Informatics Problems,
Federal Research Center “Computer Science and Control”
of the Russian Academy of Sciences, Moscow, Russia
Vavilova str. 44/2, Moscow, 119333, Russia*

*† Faculty of Computational Mathematics and Cybernetics,
Lomonosov Moscow State University
GSP-1, Leninskie Gory, Moscow, 119991, Russia*

Abstract. The paper presents a methodology for the stochastic analysis of random processes based on the method of moving separation of finite normal mixtures. Within the framework of our approach, the one-dimensional distributions of observed processes are approximated by finite location-scale mixtures of normal distributions. The theoretical background of these models is based on that finite normal mixtures are convenient approximations to general location-scale normal mixtures or normal variance-mean mixtures which are limit laws for the distributions of sums of a random number of independent random variables or non-homogeneous and non-stationary random walks and hence, are reasonable asymptotic approximations to the statistical regularities in observed real processes. We show that this approach can be also applied to positive time series, if the initial data is primarily noised by adding i.i.d. normal random variables with known parameters. This approach allows to analyze the regularities in the variation of the parameters and capturing the low-term variability in the case of complex internal structure of data.

Keywords: noisy data, moving separation of mixtures, finite normal mixtures, information systems.

1. Introduction

One of the most important indicator in the characterization of the fine structure of the processes in various modern information systems is the intensity of events, traffic, etc (see, for example, papers [1, 2]). The values of the intensity are positive, so the results of the classical statistical techniques based on the family of normal distributions can be mistaken. The paper presents a methodology for the stochastic analysis of random processes based on the method of moving separation of finite normal mixtures [3] which can be applied for positive time series correctly.

The key idea of the methodology is based on the noising of the initial data by adding independent and identically-distributed (i.i.d.) normal random variables with known parameters to refine the output. The one-dimensional distribution of the noisy sample is approximated by finite location-scale mixtures of normal distributions. It is well-known that finite normal mixtures are convenient approximations to general location-scale

normal mixtures or normal variance-mean mixtures which are limit laws for the distributions of sums of a random number of independent random variables or non-homogeneous and non-stationary random walks and hence, are reasonable asymptotic approximations to the statistical regularities in observed real processes [4].

The similar approach with the noise benefit is known in the statistical signal processing as a stochastic resonance [5–8]. The base numerical method for finding values of the unknown parameters of the model in the suggested methodology is the EM algorithm [9]. It is known [10] that the noising increases the average convergence speed of the EM algorithm.

The paper demonstrates the way to improve the quality of the structural analysis of the unknown processes in the real information systems. The approach allows to analyze the regularities in the variation of the parameters and capturing the low-term variability in the case of complex internal structure of data. An implementation of the methodology is shown by an example of the intensity in the special information system.

2. Approximation of the initial data

To analyze the changes in the mode of stochastic process, the problem of statistical estimation of unknown parameters of the model should be solved for a moving sample segment (which is called a window) of a fixed length forming the sub-samples to be further analysed. Estimating parameters for the windows, one can derive the time series of these parameters. Resulting time series of the parameters will allow for the analysis of temporal changes in the behavior of the so-called diffusive and the dynamical components in the process.

Assume that the cumulative density function for a given window centered at the time moment t can be represented as

$$F_t(x) = \sum_{i=1}^k \frac{p_i(t)}{\sigma_i(t)\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{(t-a_i(t))^2}{2\sigma_i^2(t)}\right\} dt, \quad (1)$$

where

$$\sum_{i=1}^k p_i(t) = 1, \quad p_i(t) \geq 0 \quad (2)$$

for all $x \in \mathbb{R}$, $a_i(t) \in \mathbb{R}$, $\sigma_i(t) > 0$, $i = 1, \dots, k$.

The model (1) is called a finite location-scale normal mixture. The parameters $a_1(t), \dots, a_k(t)$ are associated with the dynamic component of the internal variability of the process, and the parameters $\sigma_1(t), \dots, \sigma_k(t)$ are associated with the diffusive one, see [3]. The parameters $p_1(t), \dots, p_k(t)$ are the weights satisfying condition (2). The parameter k is the number

of mixture components and it may be also treated as depending on time. When the parameters of the model (1) are estimated for the moving segments, some weights may be very close to zero or to be evaluated as zeroes. This implies the corresponding component to vanish and the the number of components to decrease.

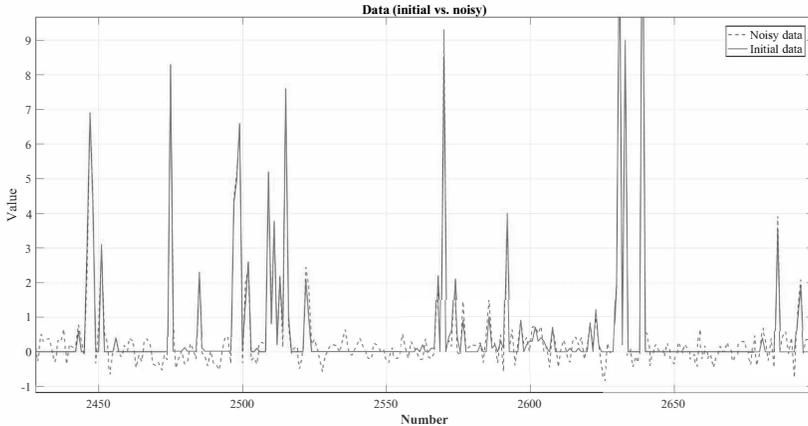


Figure 1. The comparison of the noisy data with the initial data.

The values of intensities are positive (see the solid line on the Fig. 1), so the results for the normal mixtures with the support \mathbb{R} can be mistaken. For solving this problem the initial sample should be noised by adding i.i.d. normal random variables with known parameters.

First of all, consider the approximation by normal mixtures for the initial sample. The corresponding dynamic and diffusive components are represented on Fig. 2 and 3. The x-axis for both graphs is a number of a window, it describes the location of the window in the sample. The y-axis demonstrates the values of expectation (Fig. 2) and standard deviation (Fig. 3) varying within a window moving. The color bar in the right side of the figures demonstrates the weight scale for each of components. The size of points on the figures is proportional to the weights too.

There is the one dominating component on each of the Fig. 2 and 3. The points beyond it can be interpreted as the computational errors of the EM algorithm. The statistical structure of the process cannot be analyzed in this trivial situation. Moreover, it is complicated to represent the set of points on the Fig. 3 as the curve (a diffusive component).

So, the arised problems for the analysis of the initial sample is demonstrated. In the following section we describe a procedure of the noising and compare the results for the initial and modified data.

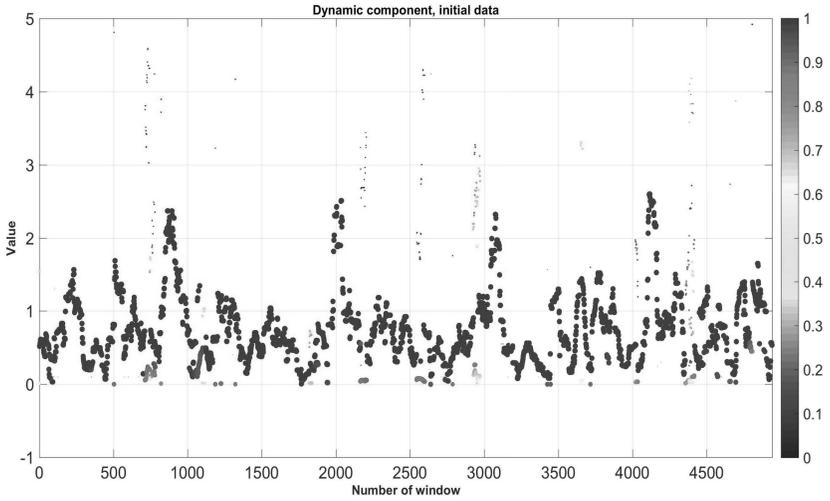


Figure 2. The dynamic component of the initial data.

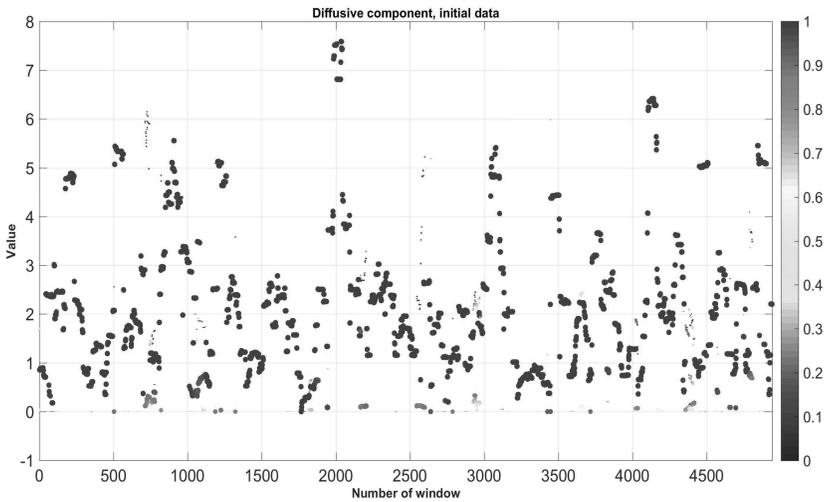


Figure 3. The diffusive component of the initial data.

3. Noising of data

Suppose that the cumulative density function of each observation X_j can be represented in form (1), that is

$$X_j \sim \sum_{i=1}^k p_i \Phi \left(\frac{x - a_i}{\sigma_i} \right)$$

The noising implies the following replacement of the original observations:

$$X_j \rightarrow X_j + \varepsilon_j,$$

for all $j = 1, \dots, N$, where N means a sample size, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ (a normal distribution with expected value 0 and standard deviation σ). As the example we use noising with the value σ equals 1% of the sample standard deviation. The choice of the value for the σ is a difficult problem due to the necessity to keep initial stochastic structure of the data. The noisy data is shown on the Fig. 1 by the dashed line.

The peak values of the noisy data are closed to the initial ones, but there are negative observations in the new sample. So, the technique, based on the normal mixtures, is correct. It should be noted that the Fig. 1 demonstrates the enlarged part of the observed samples.

In terms of the finite normal mixtures (1) the noising represents the adding of the new component with the known expectation and variance and unknown weight. It is known, that the variance for the finite normal mixtures (1) can be represented in the following form:

$$\sum_{i=1}^k p_i(t) \left[a_i(t) - \sum_{i=1}^k p_i(t) a_i(t) \right]^2 + \sum_{i=1}^k p_i(t) \sigma_i^2(t). \quad (3)$$

The noising by the $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ does not change the first summand in the expression (3), but the second one has the following form (due to the independence of the corresponding random variables $\mathbb{D}(X_j + \varepsilon_j) = \sigma_j^2 + \sigma^2$):

$$\sum_{i=1}^k p_i(t) (\sigma_i^2(t) + \sigma^2) = \sum_{i=1}^k p_i(t) \sigma_i^2(t) + \sigma^2.$$

So, it can be simply removed from the resulting approximating mixture (it should be removed from the diffusive component while the dynamic one does not change). It is visually demonstrated on Fig. 4 and 5.

The “solid” line on the Fig. 4 and 5 corresponds to the noisy component. Due to the computational errors, the expectation close to the 0 on the Fig. 4 (but it is not equals exactly). It is the estimation of the noisy standard deviation σ .

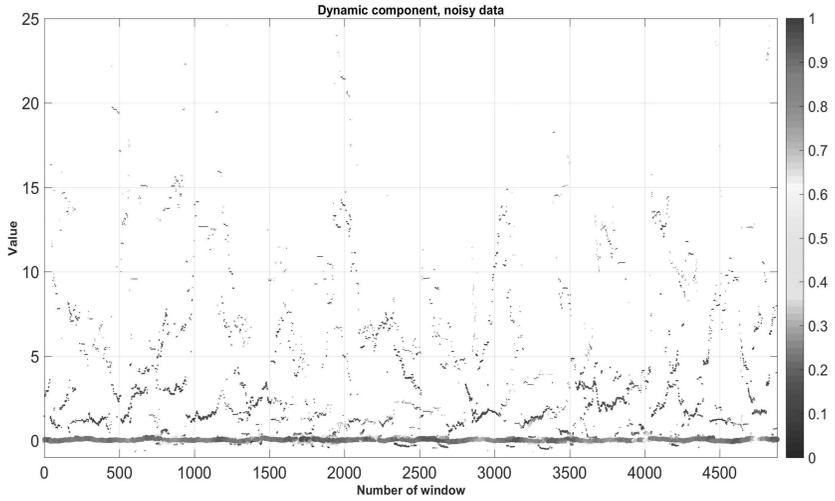


Figure 4. The dynamic component of the noisy data.

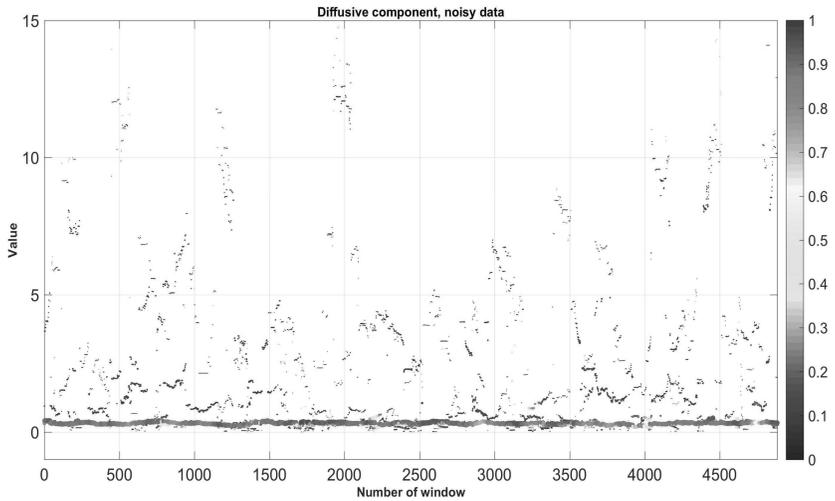


Figure 5. The diffusive component of the noisy data.

The thin lines describe the fine stochastic structure of the initial process. These results can be used for the analysis. Note that the settings of the computational procedure for the initial and noisy sample are the same: width of the window equals 120 observations, the computational accuracy is 10^{-8} , the number of components in the mixture equals 3.

4. Conclusions

The paper empirically demonstrates the efficiency of the suggested noising methodology to refine the output of the method of moving separation of finite normal mixtures. The key problem of the further research is a formulation of the convenient conditions for the parameter σ of the noisy component. The results of the paper [10] are difficult for using in practice. One of the possible solutions can be based on some information criteria (for example, the Akaike information criterion [11]) and Kullback–Leibler divergence [12] to obtain the corresponding conditions.

Acknowledgments

The research is supported by the Russian Foundation for Basic Research (projects 15-37-20851, 15-07-04040, 15-07-05316, 16-07-00736).

References

1. *Korolev V. Yu., Chertok A. V., Korchagin A. Yu., Gorshenin A. K.* Probability and statistical modeling of information flows in complex financial systems based on high-frequency data // Informatics and Applications. — 2013. — Vol. 7, no. 1. — P. 12–21.
2. *Gorshenin A. K., Korolev V. Yu., Zeifman A. I., Shorgin S. Ya., Chertok A. V., Evstafyev A. I., Korchagin A. Yu.* Modelling stock order flows with non-homogeneous intensities from high-frequency data // AIP Conference Proceedings. — 2013. — Vol. 1558. — P. 2394–2397.
3. *Korolev V. Yu.* Probabilistic and statistical methods of decomposition of volatility of chaotic processes — Moscow University Publishing House, 2011.
4. *Korolev V. Yu.* Generalized hyperbolic laws as limit distributions for random sums // Theory Probability and Its Applications. — 2014. — Vol. 58, no. 1. — P. 63–75.
5. *Brey J. J., Prados A.* Stochastic resonance in a one-dimension ising model // Physics Letters A. — 1996. — Vol. 216. — P. 240–246.
6. *Bulsara A. R., Gammaitoni L.* Tuning in to noise // Physics Today. — 1996. — P. 39–45.

7. *Gammaitoni L., Hänggi P., Jung .P, Marchesoni F.* Stochastic resonance // *Reviews of Modern Physics.* — 1998. — Vol. 70. — P. 223–287.
8. *Kosko B., Mitaim S.* Stochastic resonance in noisy threshold neurons // *Neural Networks.* — 2003. — Vol. 16, no. 5. — P. 755–761.
9. *Dempster A., Laird N., Rubin D.* Maximum likelihood from incomplete data via the EM algorithm // *Journal of the Royal Statistical Society. Series B.* — 1977. — Vol. 39, no. 1. — P. 1–38.
10. *Osoba O., Mitaim S., Kosko B.* The noisy Expectation–Maximization algorithm // *Fluctuation and Noise Letters.* — 2013. — Vol. 12, no. 3. — 1350012.
11. *Akaike H.* A new look at the statistical model identification // *IEEE Transactions on Automatic Control.* — 2013. — Vol. 19, no. 6. — P. 716–723.
12. *Kullback S., Leibler R. A.* On information and sufficiency // *The Annals of Mathematical Statistics.* — 1951. — Vol. 22, no. 1. — P. 79–86.

UDC 519.246:621.396

Computer Simulation of Average Channel Access Delay in Cognitive Radio Network

A. Yu. Grebeshkov*, A. V. Zuev*, D. S. Kiporov†

* *Chair of Automatic Telecommunications, Povolzhskiy State University of Telecommunications and Leo Tolstoy str. 6, Samara, 443010, Russia*

† *Bachelor of Department of Applied Mathematics, Samara University Samara, Moskovskoye sh. 34, Samara, 443086, Russia*

Abstract. Cognitive radio (CR) is a new wireless communication concept of the future networks, that can help to use all available radio resources at a local area. Cognitive radio is based on the dynamic spectrum access, (DSA), where available spectrum segments are used in an intelligent manner with help of advanced spectrum analysis and probing for unoccupied radio frequencies. An implementation of the cognitive radio networks raises an issue of the medium access control (MAC) protocol researching, in particular MAC protocol impacts on the access delay to radio channels. In this paper uncoordinated access method is studied where the event of spectrum and channel accessing is random and determined by probabilistic value from 0,1 to 0,99 named as channel availability. The subject of research was impact of channel availability on the access delay with simulation on the base ns2 program simulator with CRNC patch.

Keywords: cognitive radio, radio terminal device, software-defined radio, dynamic spectrum access, media access control protocol, simulator ns2.

1. Introduction

Cognitive radio (CR) is a new wireless communication concept of the future networks, that can help to use all available radio resources at a local area [1,2,3]. Cognitive radio has a clear ability to be concerning as self-configurable platform including set of different software and hardware.

Cognitive radio is based on the dynamic spectrum access, DSA. This technology explores an opportunistic spectrum access, where available spectrum segments are used in an intelligent manner with help of advanced spectrum analysis and probing for unoccupied radio frequencies. CR and DSA forms a new paradigm for radio spectrum and radio channel access and a great challenge for the traditional radio spectrum using. Regular radio access technologies (2G, 3G, 4G/LTE, WiFi) was designed on the base of centralized principle of spectrum allocation.

Now this scheme has a drawback in term of flexibility and adaptability which are the important points of advantage of cognitive radio. There are two types of users sharing a common spectrum under DSA rules:

- Primary (licensed) users who have high-priority in spectrum access and utilization within the predefined frequency bands.

- Secondary users who must access the spectrum with DSA technologies for a limited time.

The subject of the research is secondary users' character like as average channel access time delay. There is a really situation on practice when the part of frequency band previously licensed for primary user, is not being utilized by this kind of user for a short period of time. In traditional wireless systems there are no technologies that can help to use this unexpected 'white spaces' for data transmission or receiving. Cognitive radio technologies with software defined radios (SDR) are more adaptable. But with DSA there is a time period when SDR tried to access this temporary not-in-use channels for signal transmission and packet exchange. The packet transmission delay between origin and destination terminal is restricted by quality of service requirements. The average channel access time delay is the part of summarized transmission delay. In this report the DSA's access delay is under investigation with computer simulation.

2. Research issue in CR data link layers

The subject of research will be case with uncoordinated access as a more common point in CR system. The idea is that the estimation of the average access delay [4] for uncoordinated manner of access will be an upper estimation of access delay when secondary users have to wait for a time to get access to the spectrum and radio channel resource. The subject of further research with computer modelling will be a Media Access Control (MAC) protocol unit [5, 6].

The issue of best acceptable channels selection is discovered. This problem solution in term of open system interconnection (OSI) model is at the data-link level where is a control access protocol that grants an access to the transmission medium. As it said above, it is a medium access control protocol with specificity in the context of cognitive network and DSA.

MAC protocol operation bases on received data from the physical layer. This data used to solve the problem of recognition of temporary unoccupied radio frequencies. The next step is how to get access to the unoccupied channel. However, information from data link layer help to find the optimal direction / transmission path, indicating a list of available channels for the network layer. In return, the network layer can transfer to the link layer an information about which channel has an appropriable quality of service (QoS) for the data transferring session initiation.

The MAC protocol for dynamic spectrum access supports following main functions:

- The control and prevention of interference
- Prevention of conflicts of access to the channel.
- Realization of the selection process and finding unoccupied radio channel.

The computer model of dynamic access method to the cognitive network based on the Cognitive Radio Cognitive Network (CRCN) patch for network simulator ns2 on the base of Linux Ubuntu 10.04 operating system [7,8] where for CR network some special features of MAC protocol are added.

3. CRNC features for simulation

The CRCN patch has the input data as the amount of radio terminal devices with SDR features and the overall number of radio channels. The main modelling scenario provides a description of queues and channels for each SDR scenario with help of the TSL program library. Finally, the network simulator ns2 with CRCN patch has the following functionality for simulation cognitive MAC protocol unit:

1. The description of multi-channel data transmission medium.
2. The interface description for radio channel selecting.
3. A possibility to change and choice transmission power value.
4. Interference information.
5. Information about the motion or position of radio terminal devices.

The CRCN patch has settings to describe the collisions on the MAC layer. In fact, the choice of unoccupied channel depends on the MAC layer cognitive radio. Next, in the program code to the "sendDown" procedure the function "WirelessPhy" be added. This function includes the description of the frame transmission process when the frame transmits to the physical layer.

To avoid access conflicts or to reduce interference between adjacent nodes the special channel index is used in CRNC. This channel index obtained from the MAC layer or from DSA algorithm.

We are considering the model of "hello" packets transition and the routing of these messages at the same time on several radio interfaces of terminal with aim to establish communication session. A particular channel can be assigned by means of simulator to specific interface radio terminal device with SDR features. The assignation may be carried out by MAC level or can be transmitted from the network level. At the network layer the routing protocol named as *Ad hoc on-demand distance vector* (AODV) is used by procedure of routing simulation.

The test carried out on multi-channel static network structure. The aim of test is verification of network operability with procedure network layer sends available channel data to the data link layer. In the routing procedure, the algorithm AODV making decisions on the use of the assigned channel. Since management of multi-channel structure is performed by simulator for MAC layer, the test includes two stages. In the first stage, each node will send a packet to the upper OSI layer and provides with information about the unoccupied channel(s). In the second stage, the node will use selected channel to transmit and receive data.

Table 1

Simulation CRNC parameters

Description	Value
Simulation tool	NS-2 (CRNC patch)
Network area	100m x 100m
Number of nodes	20
Number of channels	1, 3, 5
Channel availability (availability)	0,1 to 0,99
Simulation time	50s.

Evaluating the probability of channel availability was described in [9]. In [10] a special monitoring network was proposed for classification of the channel availability.

Three options of multichannel structure are performed for simulation. At the first option there was 1 channel, at the second option there were 3 channels and in the end there were a 5 channels. An overall number of nodes was 20.

4. Analysis of computer simulation results

The simulation results do not provide accurate values because modeled by a random processes. In order to estimate probability p of event, where event is a case of channel occupation, the probing simulation was done with result in 14 success attempts of channel occupy during 50 tests, since probing $p = 0,28$. The accuracy evaluation (closeness in estimation) in all simulation experiments was set to $\epsilon < 0.01$. With the 95% confidence interval for the 30 points used for plot composition, the number of tests the statistical was determined as 9939, round to 10000. During simulation process was realized three experiments with 10000 tests in one experiment for 20 nodes.

The nodes placed static in random order. These nodes selected randomly as senders or receivers data. The queue service time of each network described by the exponential distribution. The node selection is randomly. In addition, necessary to note that availability defined as the probability that a channel is available for the secondary user as result of sensing and probing process.

The process of the network model includes the creation of topology and the interaction sites. It is necessary to form the grid coordinates and the size of the model. The code shows the initial stages of the interaction of components as on the Fig.1.

In the Fig. 1, tcp connection is created. The size of the transmitted packets is exposed to 2048 kbyte. As example, the node 0 sends a welcome

```

set tcp_(0) [$ns_ create-connection TCP $node_(0) TCPSink $node_(1) 0]
$tcp_(0) set window_ 32
$tcp_(0) set packetSize_ 2048
set ftp_(0) [$tcp_(0) attach-source FTP]
$ns_ at 2.5568388786897245 "$ftp_(0) start"

```

Figure 1. The code of interaction nodes on the transport layer

```

set val(chan) WirelessChannel ;#Channel Type
set val(prop) TwoRayGround ;#Radio propagation model
set val(netif) WirelessPhy ;#Network interface type
set val(ant) OmniAntenna ;#Antenna model
set val(rp) AODV ;#Routing Protocol
set val(ifq) Queue/DropTail/PriQueue ;# interface queue type
set val(ifqlen) 500 ;# max packet in ifq
set val(mac) Mac/Macng ;# MAC type
set val(ll) LL ;# link layer type
set val(nn) 20 ;# number of mobilenodes
set val(channum) 1/3/5 ;# number of channels per radio
set val(cp) ./topo.tcl ;# topology traffic file
set val(stop) 50 ;# simulation time

```

Figure 2. Base settings of the computer simulation model

message to node 1. The more detailed description of simulation model is at the Fig.2.

The network layer transmits routing information about the available channel to the lower level. MAC-level is available in the multichannel structure.

On the Fig. 3 the number of the channels sets as "C".

Simulation is carried out with a different number of channels ($C = 1, 3, 5$), nodes get access to the channel at the same time, this procedure leads to access delay and collisions in the simulated network.

When availability value is observed from 0,1 to 0,25 approximately, there is high-delay access delay (from 15 ms to 25 ms), but when the availability is increasing the number of collisions is reduced. The matter of the observation is visible crossing points of lines on the plot in Fig. 3. On the plot's right side from the crossing point (for availability 0,6) is not too much changing of the channel access delay (delay changing from 10 ms to 12 ms) for one-channel scheme. But in interval of availability value from 0,6 to 1 there is delay changing from 9 ms (approximately) to 2 ms for 3-channel and from 5,5 ms to 2 ms for 5-channel scheme. Since in real network there is a possibility with specific value of availability emerging to decrease number of accessible channels without greater increasing of average channel access delay. This result could be used in radio devices tune-up and for network design, planning and analysis. The next conclusion is one-channel scheme could be considered as partial stability in average access delay in some part of availability value (from 0,3 to 1) as shown in Fig. 3.

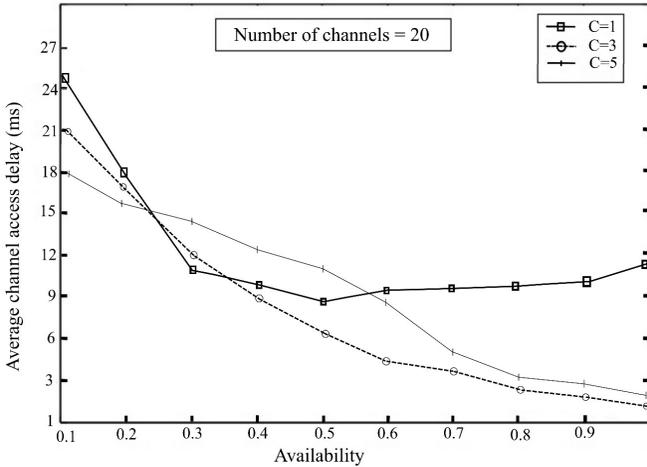


Figure 3. Average channel access delay by using 20 nodes

It is need to remark there are some rises and falls in the plot at the Fig. 3 because the process of channel selecting is a random and the channel availability is random value too.

The results of simulation show there is a threshold value availability for static ad-hoc network in context of average channel access delay. If the value availability will be bigger than threshold value (approximately 0,6-0,7) than 3-channels network scheme looks like 5-channels network scheme and one-channel scheme has not dramatically increasing of delay (for availability between 0,3 and 0,6) as an effect of cognitive network features.

In the future DSA with cognitive radio principles and software configurable radio-terminals create a wide range opportunities for reconfigurable radio networks in military, social works, medicine, radio access technology new generation including cognitive wireless sensor networks.

5. Conclusions

Cognitive radio principles create a wide range opportunities for research and applications in military, social works, medicine, radio access technology new generation. The great aim of these technologies is to improve the spectrum efficiency using for radio terminal devices with SDR features.

The actual issue is the research of future parameters estimation of the cognitive radio networks, like as average channel access delay for secondary users. The results of computer simulation for uncoordinated access method

shows that access delay value depends of availability of radio channels. There is a possibility to decrease number of accessible channels without dramatically average channel access delay changing at the network with fixed nodes. For the further studying, the problem of the access delay for coordinated access method and network with high-rate nodes will be important.

References

1. *Rao K. R., Bojkovic B. M.* B.M. Wireless multimedia communication systems. Design, Analysis and Implementation. — 2014.
2. *Oiu R. C., Hu Z., Li H., Wicks M.C* Cognitive radio communications and networking. Principle and practice. — Wiley, 2012.
3. *Zhang Y., Zheng J.* Cognitive radio networks: architectures, protocols, and standards. — CRC Press, 2010.
4. *Report ITU-R SM.2256.* Spectrum occupancy measurement and evaluation. SM series. Spectrum management, ITU electronic publication, Geneva, 2012.
5. *De Dominico A., Calvanese S. E.* A survey on MAC strategies for cognitive radio networks. IEEE Comm. surveys & tutorials. — Vol. 14, iss. 1, 2012. — P.21-44.
6. *Shahid Khan U., Maqsood T.* CRN survey and a simple sequential MAC protocols for CRN learning // Second international conference on advances in cognitive radio, COCOR 2012 / Concordia university, Canada. — P. 22-27, 2012.
7. *Lee P., Wey g.* NS2 model for cognitive radio networks routing // Proceedings of the First International Symposium on CNMT 2009 (Computer Network and Multimedia Technology) / Wuhan, China. — P. 1-4, 2009.
8. *Bhrugubanda M.* A Survey on simulators for cognitive radio network. International journal of computer science and information technologies IJCSIT. — Vol. 5(3). — P.4760-4761, 2014.
9. *Kanizhil R., Chandrasekar C.* Evaluating the probability of channel availability for spectrum sharing using cognitive radio. International journal of engineering research and applications (IJERA). — Vol. 2, iss 4. — P.2186-2197, 2012.
10. *Canberk B., Oktug S.* A channel availability classification for cognitive radio networks using a monitoring network // 17th IEEE ISCCS2012 (Symposium on Computers and Communications) / P. 690-695, 2012.

УДК 621.395

Анализ влияния технологий D2D на функционирование беспроводных сетей связи

О. А. Хуссейн*, Д. В. Окунева†

* *Кафедра Сетей связи и передачи данных, Санкт-Петербургский государственный университет телекоммуникаций им. проф.*

М.А.Бонч-Бруевича,

пр.Большевиков д.22, корп.1, Санкт-Петербург, Россия, 193232

† *Кафедра программной инженерии и вычислительной техники, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича,*

пр.Большевиков д.22, корп.1, Санкт-Петербург, Россия, 193232

Аннотация. Научные исследования в области построения сетей 5G являются приоритетными и востребованными в мире, так как направлены на подготовку к очередному шагу развития сетей подвижной связи. Одной из особенностей современной концепции построения сетей 5G является использование технологий D2D, позволяющих в ряде случаев существенно влиять на устойчивость и качество предоставления услуг связи, например, таких как услуги дополненной реальности. Статья содержит результаты анализа задач построения сетей 5G. Рассмотрены возможности применения технологий D2D коммуникаций, особенности реализации технологии D2D в сетях пятого поколения, результаты анализа эффекта от использования технологий D2D, анализ изменения доступности терминала и изменения трафика на элементы сети.

Ключевые слова: 5G - пятое поколение, D2D – устройство-устройство, повышение доступности, снижение трафика, связность сети.

1. Введение

Эволюция технологий беспроводной связи привела к широчайшему проникновению услуг беспроводной радиотелефонной связи и беспроводного доступа. Рост пропускной способности на уровне абонентского доступа и плотности абонентов, внедрение новых услуг приводят к существенному росту интенсивности абонентского трафика. Это в ряде случаев приводит к снижению качества обслуживания на уровне доступа. Для предотвращения этого нежелательного явления приходится прибегать к уменьшению зоны обслуживания базовых станций и увеличению их количества [4]. В основе построения современных сетей беспроводной связи лежит принцип многогранговой архитектуры, на нижнем уровне которой находятся абонентские терминалы, а на следующем узлы доступа – базовые станции или точки доступа. Данный принцип исключает возможность горизонтальных связей (терминал – терминал). Это позволяет полностью контролировать функционирование абонентских терминалов, однако исключает возможность их непосредственного взаимодействия и требует покрытия всей области

их расположения базовыми станциями. Следует отметить, что взаимодействие терминалов без участия базовых станций реализовано, например, в системах радиосвязи точка-точка и системах транкинговой связи, однако, из-за специфики их применения оно получило ограниченное распространение. Непосредственное взаимодействие между терминалами (далее D2D) позволяет повысить такие показатели сети как надежность и устойчивость к различным деструктивным факторам, в том числе снизить интенсивность абонентской нагрузки на базовые станции. В ряде случаев, например, в условиях чрезвычайных ситуаций применение технологий D2D позволяет значительно увеличить эффективность связи и управления. Развитие принципов построения одноранговых самоорганизующихся сетей [5] явилось стимулом их реализации в концепции построения перспективных сетей пятого поколения (5G) наряду с традиционными принципами построения многогранговой сети. В данной работе приводятся результаты анализа влияния технологий D2D на основные показатели функционирования сети связи.

2. Особенности реализации технологии D2D в сетях пятого поколения

Технология D2D в сетях 5G предполагает возможность организации соединений типа:

-Терминал А - терминал R (или несколько терминалов) – БС. Устройства взаимодействуют с базовой станцией через ретрансляцию информации с помощью других устройств.

-Терминал А - БС - терминал В. Имеет место прямое взаимодействие без участия базовой станции, Но их взаимодействие координируется оператором.

-Терминал А - терминал R (или несколько терминалов) - терминал В. Устройства источника и получателя несут ответственность за координацию взаимодействия с использованием ретрансляторов. В этом случае устройства источника и потребителя имеют прямую связь друг с другом без какого-либо контроля со стороны оператора,

-Терминал А – терминал В. имеют прямую связь друг с другом без участия базовой станции и без какого-либо контроля со стороны оператора (рис.1).

Возможность установления связи D2D определяется возможностью организации маршрута между терминалами А и В. Эта возможность определяется взаимным расположением этих терминалов и терминалов, которые могут выполнять функции транзита трафика. Таким образом, возможность организации D2D коммуникаций в значительной степени определяется распределением пользователей сети.

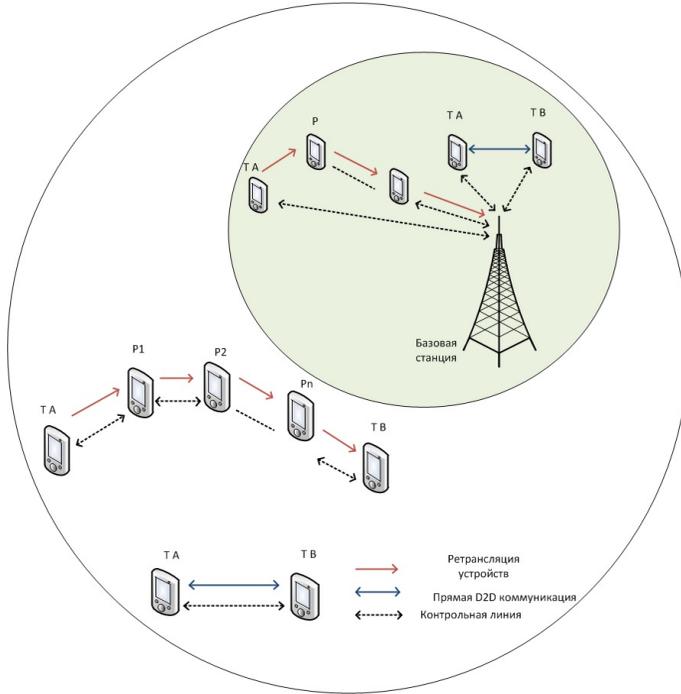


Рис. 1. Варианты коммуникаций в сети 5G

3. Эффект от использования технологий D2D

Повышение доступности. Будем полагать, что терминалы сети распределены в зоне обслуживания случайным образом и образуют пуассоновское поле [1]. При достаточно большом числе терминалов можно воспользоваться моделью случайного графа [3]. Тогда, следуя модели Эрдеша–Реньи [2] вероятность доступности терминала B терминалу A будет определяться связностью сети, которая зависит от вероятности существования ребра графа (связи) между двумя произвольными узлами сети i и j . Пороговая величина этой вероятности определяется как $p = \ln(n)/n$, где n число терминалов сети. Данная величина характеризует так называемое состояние фазового перехода сети от несвязного состояния сети к связному и наоборот. При уменьшении $p < \ln(n)/n$ сеть переходит в несвязное состояние, а при увеличении $p > \ln(n)/n$ в связное. Вероятность p для терминала сети определяется средним числом доступных ему терминалов. Для пуассоновского поля она равна $\bar{k} = \rho S$, где ρ – плотность терминалов, S – площадь

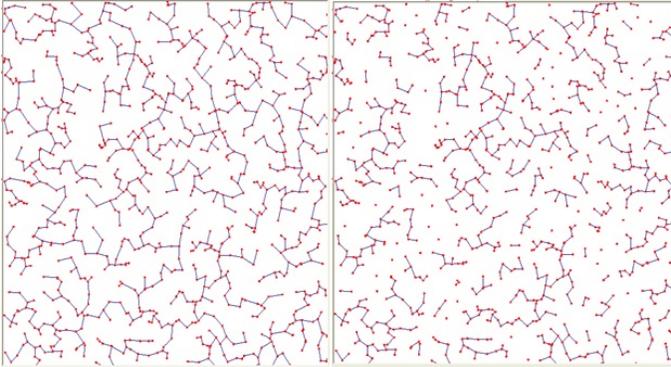


Рис. 2. Зависимость связности сети от радиуса связи терминала

зоны связи терминала. На рис.2 приведены результаты имитационного моделирования связности сети, и ее зависимости от радиуса связи терминала. При уменьшении радиуса связи до 20 м приведенная структура сети «рассыпается» на изолированные подсети (кластеры) и изолированные терминалы. Теоретически, при достаточно большом радиусе связи, который можно вычислить из определения пороговой вероятности и свойств пуассоновского поля, вероятность связности терминалов A и B может быть достаточно близка к единице.

В этом случае не ограничивается число возможных транзитов, что на практике недопустимо как из-за сложности реализации процедур маршрутизации, так и из-за снижения качества обслуживания. Приведенный пример можно сопоставить с распределением абонентов в объектах городской инфраструктуры: жилых и иных зданиях и сооружениях, на тротуарах и в транспорте. Эти элементы представляют собой области компактного размещения пользователей, в пределах которых радиус связи терминала достаточен для обеспечения связности сети в границах этого элемента. Из условий связности можно определить соотношение между радиусом связи терминала и плотностью пользователей, при котором достигается необходимая вероятность связности сети. Рассматривая 3D модель при случайном распределении терминалов в пространстве (пуассоновское поле) можно оценить вероятность доступности терминала как

$$\bar{p} = \frac{4\pi R^3}{3n} \rho \quad (1)$$

где R – радиус связи терминала (м), n – количество терминалов в рассматриваемом пространстве, ρ – плотность терминалов (шт./м³). С учетом (1) условие связности сети из пороговой вероятности определим как

$$\bar{p} > p \Rightarrow \frac{4\pi R^3}{3n} \rho > \frac{\ln(n)}{n} \Rightarrow \frac{4\pi R^3}{3} \rho > \ln(n) \Rightarrow R > \sqrt[3]{\frac{3\ln(n)}{4\pi\rho}} \quad (2)$$

Из модели Эрдеша–Реньи следует, что при выполнении условия (2) сеть переходит в связанное состояние. Например, при допущении, что плотность терминалов составляет $4,4 \cdot 10^{-3}$ терминалов/м³, а их количество равно 200, из условия (2) получаем, что радиус связи узла должен превышать 6,6 м. Эти численные значения ориентировочно описывают сеть в 200 квартирном жилом здании, при условии, что в количестве терминалов в ней равно количеству квартир. Разумеется, данный пример не учитывает особенностей планировки здания и расположения жилых помещений, однако он дает возможность оценки порядка этой величины. Если оценить возможный радиус связи внутри помещения, например, для стандарта IEEE 802.11n [6], с помощью модели ITU-R P.1238-7 [7], то он будет составлять 25...50 м, в зависимости от особенностей конструкции здания. Сопоставляя эту величину с полученным выше значением 6,6 м можно предположить, что в границах таких объектов городской инфраструктуры как жилые, офисные, производственные и другие здания может быть достигнута высокая вероятность связности D2D коммуникаций. Приведенные выше рассуждения дают основания полагать, что возможность D2D коммуникаций потенциально позволяет повысить доступность терминалов, находящихся вне зоны действия БС, а также существенно повысить устойчивость сети за счет возможности связи в локальных зонах.

Снижение трафика на БС. Возможность организации соединений D2D дает потенциальную возможность замыкания доли трафика непосредственно между пользователями, минуя инфраструктуру сети. При некоторых допущениях можно предположить, что вероятность пропуска трафика через соединение D2D равна вероятности связности. Тогда трафик производимый пользователями сети будет распределяться на инфраструктуру сети и D2D коммуникации согласно выражению

$$y = y_{BS} + y_{D2D}$$

где y_{BS} —трафик пользователей, производимый на инфраструктуру сети, y_{D2D} —трафик пользователей, обслуживаемый соединениями D2D.

$$y_{D2D} = p_c y$$

$$y_{BS} = (1 - p_c) y$$

где p_c – вероятность связности. Согласно приведенной в выше модели вероятность связности можно охарактеризовать с помощью пороговой вероятности доступности $p_0 = \ln(n)/n$.

Если предположить гипотетическую 2D модель сети города с 4 млн. абонентов (терминалов) равномерно распределенных по площади 1439 км², то абонентская плотность составит $2,78 \cdot 10^{-3}$ терминалов/м². При радиусе связи терминала $R=50$ м доступность составит $5,46 \cdot 10^{-6}$, в то время как $p_0 \approx 3,8 \cdot 10^{-6}$. Следовательно, при данных условиях можно предположить, что сеть является связной. Данная модель не учитывает неравномерности распределения абонентов по территории города, многоквартирных зданий, акваторий, а также ограничений на длину маршрута. Однако, на этом примере можно видеть, что в городе с достаточно большим числом абонентов имеется потенциальная возможность обслуживания значительной доли трафика соединениями D2D.

4. Выводы

В работе были рассчитаны гипотетические модели связности между терминалами в жилом доме и на территории, эквивалентной по площади и числу терминалов многонаселенному городу. Полученные результаты показали высокую вероятность связности в том и другом случае. Основные результаты показали, что технология D2D в сетях 5G:

1. Позволяет существенно расширить возможности по взаимодействию между пользователями сети.

2. Позволяет повысить доступность терминалов, находящихся вне зоны действия БС, а также существенно повысить устойчивость сети к деструктивным воздействиям.

3. В городе с достаточно большим числом абонентов имеется потенциальная возможность обслуживания значительной доли трафика соединениями D2D.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00209 мол а "Разработка принципов интеграции технологии дополненной реальности и Интернета Вещей."

Литература

1. *Вентцель Е.С.* Теория вероятностей. М.: «Наука», 1969 г. 576 с.
2. *Дистель Р.* Теория графов. Новосибирск, «Издательство института математики», 2002 г. 335 с.
3. *Райгородский А.М.* Модели случайных графов и их применения. ТРУДЫ МФТИ. — 2010. — Том 2, № 4. С. 300-325.

4. *Tehrani M.N.* Device-to-Device Communication in 5G Cellular Networks: Challenges, Solutions, and Future Directions. // IEEE Communications Magazine . / M.N. Tehrani, M. Uysal, H. Yanikomeroglu. May, 2014 IEEE. PP 86-92.
5. *Yilmaz O. N. C.* Uusitalo Smart mobility management for D2D communications in 5G networks. // Wireless Communications and Networking Conference Workshops (WCNCW). / O. N. C. Yilmaz, Zexian Li ; K. Valkealahti ; M. A., 2014 IEEE. PP 219-223.
6. IEEE 802.11n – 2009 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification. P. 502.
7. ITU-R P.1238-7 Propagation data and prediction methods for the planning of indoor radio communication systems and the radio local area networks in the frequency range 900 MHz to 100 GHz. Geneva: ITU-R Recommendations, 2001. P. 22

UDC 621.395

Analysis of D2D technologies impact on the operation of wireless networks

O. A. Hussein*, D. V. Okuneva*

** Federal State Budget-Financed Educational Institution of Higher Education
The Bonch-Bruевич Saint - Petersburg State University of Telecommunications
22-1, Prospekt Bolshevikov, 193232, Saint-Petersburg, RUSSIA*

Scientific research in the field of construction of 5G networks are a priority and demand in the world, as are aimed at preparation for the next step in the development of mobile networks. One of the features of the modern concept of building 5G networks is the use of D2D technologies, allowing in some cases significantly affect the stability and quality of communication services. article provides an analysis of 5G networks build tasks. The possibilities of use of D2D communications technologies, particularly the implementation of D2D technology in the fifth-generation networks, the results of the analysis of the use of technology D2D effect, analysis of changes in the availability of the terminal and changes the traffic on the network elements.

Keywords: 5G i the fifth generation, D2D - device to device, improving accessibility, traffic reduction, network connectivity.

УДК 519.872

Исследование RQ-систем с приоритетом заявок

Я. Е. Измайлова

*Кафедра теории вероятностей и математической статистики,
Национальный исследовательский Томский государственный
университет,
пр. Ленина, д.36, Томск, Россия, 634050*

Аннотация. В данной работе исследуются RQ-системы с вытеснением заявок. Исследования проводятся с помощью метода асимптотического анализа в предельном условии большой задержки, численного анализа, диффузионной аппроксимации.

Ключевые слова: RQ-система, вытеснение заявок, асимптотический анализ.

Введение

В последнее время возродился интерес к задачам ТМО (теории массового обслуживания), обусловленный не только новыми проблемами, возникшими в практической жизни и особенно в областях, связанных с разработкой и применением вычислительной техники, но и новыми математическими подходами к их решению. Применение классических моделей теории массового обслуживания для повышения надежности прогнозирования и обработки информации телекоммуникационных, вычислительных и экономических систем не всегда дает адекватные результаты. Поэтому для анализа и исследования таких систем используют более адекватные модели, модели с повторной очередью (Retrial Queueing System). В классических RQ-системах предполагается, что все запросы, поступающие в систему - однородные, однако, во многих реальных системах запросы, поступающие в систему, неоднородны как по распределению времени обслуживания, так и по их ценности для системы и, следовательно, имеют право претендовать на первоочередное обслуживание в момент освобождения прибора. Системам с приоритетами посвящено немало исследований, к которым можно отнести работы Cobham A., Phipps, Schrage L. E., Jaiswal N. K., Madan K. C., Simon B., K. Altinkemer, I. Bose и R. Pal., Martin M., Krishna Reedy G. V., Nadarajan R, П. П. Бочарова, О. И. Павловой .

1. Исследование RQ-системы $M|GI|1$ с вытеснением альтернативных заявок

В работе исследуется RQ-система $M|GI|1$ с вытеснением альтернативных заявок и экспоненциальной задержкой заявок в ИПВ.

Для данной системы найдена пропускная способность. Сформулирована и доказана следующая теорема.

Теорема 1. Пропускная способность S RQ - системы $M|GI|1$ с вытеснением заявок имеет вид

$$S = B'(0).$$

Пропускная способность данной системы может иметь следующие значения

$$S = \begin{cases} 0, & B'(0) = 0, \\ B'(0), & 0 < B'(0) < \infty, \\ \infty, & B'(0) = \infty. \end{cases}$$

Так как условием существования стационарного режима в системе массового обслуживания с пропускной способностью S является неравенство $\lambda < S$, то для рассматриваемой RQ - системы при $B'(0) = \infty$ стационарный режим существует при любых конечных значениях интенсивности λ входящего потока. При $B'(0) = 0$ стационарного режима в данной системе не существует при любых, даже сколь угодно малых положительных значениях интенсивности λ .

Был разработан численный алгоритм нахождения двумерного стационарного распределения вероятностей числа заявок в источнике повторных вызовов и состояний прибора. Численный алгоритм содержит следующие этапы:

1. Полагаем $P_1(0) = 1$;

2. Выбираем достаточно большое целое N и полагаем, что $i = \overline{1, N}$, а $P_1(i)$ определяются равенством $P_1(i) = \frac{[\lambda + i\sigma - \lambda B^*(\lambda + i\sigma)] \frac{\lambda}{i\sigma} - \lambda B^*(\lambda + i\sigma)}{(\lambda + i\sigma) B^*(\lambda + i\sigma)}$. $P_1(i - 1)$;

3. Применяя равенство

$$P_0(0) = \frac{\sigma}{\lambda} \cdot \frac{B^*(\lambda)}{1 - B^*(\lambda)} P_0(1),$$

находим значение $P_0(0)$;

4. Применяя равенство $P_0(i + 1) = \frac{\lambda}{(i+1)\sigma} P_1(i)$, находим значения $P_0(i)$ для всех $i = \overline{1, N}$;

5. Определяем нормирующую величину $d = \sum_{i=0}^N (P_0(i) + P_1(i))$;

6. Двумерное распределение вероятностей полагаем равным $\frac{1}{d} P_k(i)$;

7. Величина N выбирается из условия, что нормированное значение вероятности $P(N)$ достаточно мало, например, равняется величине машинного нуля.

Для RQ-системы получена асимптотическая характеристическая функция вида

$$h_2^{as}(u) = \exp\left\{ju\frac{\kappa_1}{\sigma} + \frac{(ju)^2}{2}\frac{\kappa_2}{\sigma}\right\},$$

которая позволяет найти аппроксимацию второго порядка распределения $P(i)$, или будем называть ее гауссовской аппроксимацией. Также получена функция

$$h_3^{as}(u) = \exp\left\{ju\frac{\kappa_1}{\sigma} + \frac{(ju)^2}{2}\frac{\kappa_2}{\sigma} + \frac{(ju)^3}{6}\frac{\kappa_3}{\sigma}\right\},$$

которая не является характеристической, но позволяет построить дискретное распределение, которое будем называть аппроксимацией третьего порядка с моментами, найденными через асимптотические семиинварианты $\kappa_1, \kappa_2, \kappa_3$. Семиинварианты найдены в работе.

Была проведена диффузионная аппроксимация RQ-системы M|GI|1 с вытеснением альтернативных заявок и экспоненциальной задержкой заявок в ИПВ. Сформулирована и доказана следующая теорема.

Теорема 2. Распределение вероятностей $P(y, \tau)$, определяемое характеристической функцией $F_0(w, \tau, \varepsilon) + F_1(w, \infty, \tau, \varepsilon)$ при $\varepsilon \rightarrow 0$, удовлетворяет уравнению Фоккера-Планка следующего вида

$$\frac{\partial P(y, \tau)}{\partial \tau} = -\frac{\partial}{\partial y} \{a(x(\tau))yP(y, \tau)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{b(x(\tau))P(y, \tau)\},$$

где $a(x(\tau))y$ – коэффициент переноса, $a(x(\tau)) = R_1^*(x(\tau)) - R_0(x(\tau))$, а $b(x(\tau)) = x(\tau)R_0(x(\tau)) + \lambda R_1(x(\tau)) + 2x'(\tau)R_1^*(x(\tau))$ – коэффициент диффузии. Найдена плотность распределения вероятностей $\pi(z)$ значений диффузионного процесса $z(\tau) = \sigma i \left(\frac{\tau}{\sigma}\right) = x(\tau) + \sqrt{\sigma}y(\tau)$, с помощью, которой можно построить дискретное распределение. Его будем называть диффузионной аппроксимацией исходного распределения вероятностей.

2. Исследование RQ-системы $M^{(2)}|B(x)^{(2)}|1$ с g -настойчивым вытеснением альтернативных заявок

В работе также исследуется система типа $M^{(2)}|B(x)^{(2)}|1$ с вытеснением альтернативных заявок. Ставится задача нахождения совместного распределения вероятностей числа заявок в источниках повторных вызовов и состояний прибора. Показано, что асимптотическая функция имеет также вид нормального распределения

$$H(u_1, u_2) = \exp\left\{j\frac{u_1}{\sigma}x_1 + j\frac{u_2}{\sigma}x_2 + \frac{(ju_1)^2}{2\sigma}Q_{11} + \frac{(ju_2)^2}{2\sigma}Q_{22} + \frac{ju_1ju_2}{\sigma}Q_{12}\right\}.$$

Здесь x_1, x_2 являются решением системы уравнений

$$-\gamma_1 x_1 R_0(x_1, x_2) + (\lambda_1 + r_2 \lambda_2 + r_2 \gamma_2 x_2) R_1(x_1, x_2) + (-r_1 \gamma_1 x_1 + \\ + (1 - r_1) \lambda_1) R_2(x_1, x_2) = 0,$$

$$-\gamma_2 x_2 R_0(x_1, x_2) + (-r_2 \gamma_2 x_2 + (1 - r_2) \lambda_2) R_1(x_1, x_2) + (\lambda_2 + \\ + r_1 \lambda_1 + r_1 \gamma_1 x_1) R_2(x_1, x_2) = 0,$$

а $R_k(x_1, x_2), k = 0, 1, 2$ определяются равенствами

$$R_0(x_1, x_2) = 1 - \lambda_1 \frac{1 - B_1^*(r_2(\lambda_2 + \gamma_2 x_2))}{r_2(\lambda_2 + \gamma_2 x_2) B_1^*(r_2(\lambda_2 + \gamma_2 x_2))} - \\ - \lambda_2 \frac{1 - B_2^*(r_1(\lambda_1 + \gamma_1 x_1))}{r_1(\lambda_1 + \gamma_1 x_1) B_2^*(r_1(\lambda_1 + \gamma_1 x_1))}, \\ R_1(x_1, x_2) = \lambda_1 \frac{1 - B_1^*(r_2(\lambda_2 + \gamma_2 x_2))}{r_2(\lambda_2 + \gamma_2 x_2) B_1^*(r_2(\lambda_2 + \gamma_2 x_2))}, \\ R_2(x_1, x_2) = \lambda_2 \frac{1 - B_2^*(r_1(\lambda_1 + \gamma_1 x_1))}{r_1(\lambda_1 + \gamma_1 x_1) B_2^*(r_1(\lambda_1 + \gamma_1 x_1))}.$$

Литература

1. Artalejo J.R.: Accessible Bibliography on Retrial Queues: Progress in 2000-2009. Mathematical and Computer Modeling, vol. 51., pp. 1071–1081 (2010)
2. Falin G.I.: A Survey of Retrial Queues. Queueing Systems, vol. 7, pp. 127–167 (1990)
3. Falin G.I., Artalejo J.R., and Martin M.: On the single retrial queue with priority customers. Queueing Systems, vol. 14, No. 3–4, pp. 439–455 (1993)
4. Choi, B.D., Chang, Y.: Single Server Retrial Queues with Priority Calls. Mathematical and Computer Modeling vol. 30, No. 3–4, pp. 7–32 (1999)
5. Choi, B.D., Choi, K.B., Lee, Y.W.: M/G/1 retrial queueing systems with two types of calls and finite capacity. Queueing Systems, vol.19, pp.215–229 (1995)
6. Choi, B.D., Park, K.K.: The M/G/1 retrial queue with bernoulli schedule. Queueing Systems, vol. 7, No. 2, pp. 219–227 (1990)
7. Choi, B.D., Shin, Y.W., Ahn, W.C.: Retrial queues with collision arising from unslotted CSMA/CD protocol. Queueing Systems, vol. 11, No. 4, pp. 335–356 (1992)

8. Choi, B.D., Park, K.K., Pearce, C.E.M.: An M/M/1 retrial queue with control policy and general retrial times. *Queueing Systems*, vol. 14, No. 3–4, pp. 275–292 (1993)
9. Rengnanathan, N., Kalayanaraman, R. and Srinivasan, B.: A finite capacity single server retrial queue with two types of calls. *International Journal of Information and Management Sciences*, vol.13, No.3, pp.47–56 (2002)
10. Zhu, Y.J., Zhou, Z.H., Feng, Y.G.: M/G/1 retrial queue system with priority and repair. *Zidonghua Xuebao/Acta Automatica Sinica*, vol. 34, No. 2, pp. 195–201 (2008)
11. Moreno, P: An M/G/1 retrial queue with recurrent customers and general retrial times. *Applied Mathematics and Computation*, vol. 159, No. 3, pp. 651–666 (2004)
12. Bocharov, P. P., Pavlova, O. I. and Puzikova, D. A.: M|G|1|r retrial queueing systems with priority of primary customers. *Mathematical and computer Modelling*, vol. 30, No. 3–4, pp. 89–98
13. D'Apice, C., Simone, T. De., Manzo, R., Rizelian, G.: Priority Service of Primary Customers in the M/G/1/r Retrial Queueing System with Server Searching for Customers. *Informacionny processy*. vol. 4, No. 1. pp.13–23 (2004)

UDC 519.872

The research of retrial queueing systems with exclusion of customers

Y. Izmaylova

*Department of probability theory and mathematical statistics,
Tomsk State University
Russia, 634050, Tomsk, 36 Lenin Pr.*

In this paper, we research retrial queueing systems with exclusion of customers by method of asymptotic analysis with condition of long delay of customers in the orbit and method simulation modeling, numerical realization.

Keywords: retrial queueing systems, exclusion customers, asymptotic analysis.

УДК 519.87

О скорости сходимости к стационарному распределению ненадёжной сети Джексона с динамической маршрутизацией

Э. Ю. Калимулина

*Институт проблем управления
им. В. А. Трапезникова РАН
Россия, г. Москва, Профсоюзная ул., 65*

Аннотация. В работе рассматривается открытая сеть массового обслуживания типа Джексона с ненадёжными узлами. Сеть состоит из $m < \infty$ узлов, каждый узел сети - система массового обслуживания типа $M/G/1$. Предполагается, что поток заявок, поступающий в сеть - пуассоновский с параметром $\Lambda(t)$. Задана матрица маршрутизации потоков $\{r_{ij}\}$, $i, j = 0, 1, \dots, m$, $\sum_{i=1}^m r_{0i} \leq 1$. С вероятностью r_{0i} заявка направляется на обслуживание в узел i , где обслуживается с интенсивностью $\mu_i(t, n_i(t))$, зависящей как от времени t , так и от количества заявок в узле $n_i(t)$. Узлы в сети могут отказывать и восстанавливаться с некоторыми интенсивностями, зависящими от числа уже отказавших узлов. В работе предполагается, что в случае отказа некоторого узла j заявка из узла i отправляется на обслуживание в первый доступный узел, расстояние от которого до узла j минимально, т.е. рассматривается динамическая маршрутизация в случае отказа некоторых узлов. В работе приводятся результаты для оценки скорости сходимости к стационарному распределению для данной модели сети.

Ключевые слова: динамическая маршрутизация, система массового обслуживания $M/G/1$, ненадёжная сеть массового обслуживания, сеть Джексона, скорость сходимости, стационарное распределение.

1. Введение

Системы и сети массового обслуживания являются наиболее адекватным математическим аппаратом для моделирования и оценки производительности современных телекоммуникационных, транспортных и других систем [1–5]. С другой стороны, важным критерием оценки качества подобных систем является надёжность [6], поэтому большой интерес в последнее время представляют модели с ненадёжными элементами. Довольно большое число работ ограничивается исследованием систем массового обслуживания с ненадёжными приборами [7–12].

Цель данной работы - исследование математической модели сетевой системы как совокупности взаимосвязанных узлов, которые могут отказывать и восстанавливаться. В статье предлагается модификация классической модели открытой сети массового обслуживания Джексона (см., например, [13, Глава 2]), основанная на принципе динамической маршрутизации.

Для начала уточним понятие “динамическая маршрутизация”. Строгого математического определения для динамической маршрутизации нет. Отметим ряд работ, в которых рассматриваются модели сетей массового обслуживания с динамической маршрутизацией. В работах Введенской [14, 15] рассматриваются сети с фиксированной матрицей маршрутизации между узлами, каждый узел - многоканальная система, принцип динамической маршрутизации заключается в случайном выборе канала в узле. В работах F. P. Kelly (см., например [16]) сеть рассматривается как совокупность параллельных очередей, в сеть поступает несколько типов заявок и принцип динамической маршрутизации заключается в выборе той или иной очереди в зависимости от типа заявки. В данной работе под динамической маршрутизацией будет пониматься перестроение маршрута передачи сообщения в зависимости от отказа/восстановления определенного узла в сети. Идея динамической маршрутизации состоит в изменении матрицы маршрутизации потоков $\{r_{ij}\}$ и перенаправлении заявок в случае отказа узла j на доступные узлы, т.е. данное понятие здесь максимально приближено к тому, как оно определяется в телекоммуникационных сетях [18].

Отметим, что к задаче оценки надёжности сетей могут применяться и другие подходы. Для исследования связанности, например, могут использоваться модели случайных графов Эрдеша-Реньи (Erdos-Renyi graph) [19]. Но, поскольку, наша задача - оценка производительности и анализ потоков трафика в сети с учётом надёжности, то мы используем аппарат теории массового обслуживания и модель Джексона.

Математические модели ненадёжных сетей массового обслуживания неоднократно рассматривались в литературе. Отметим работу Н. Дадина [17], в которой приводится несколько вариантов модификации матрицы $\{r_{ij}\}$ в случае отказа узлов, все они ограничены тем, что основаны на принципе блокировке заявок и повторном обслуживании после восстановления. В данной работе предлагается новый вариант расчета матрицы $\{r_{ij}\}$. Среди работ, где рассматриваются модели ненадёжной сети и оценивается скорость сходимости к стационарному распределению, выделим работы Loreka и Szekli [20, 21]. Здесь будет использована техника, предложенная данными авторами для доказательства некоторых оценок, но будет рассмотрена более общая модель для узлов сети.

2. Определение процесса

2.1. Классическая модель сети Джексона

Классическая модель сети Джексона состоит из m узлов ($m < \infty$), $M := \{1, \dots, m\}$. Каждый узел - одноканальная система массового обслуживания типа $M/M/1$ с дисциплиной обслуживания в порядке поступления заявок. Заявки, поступающие в сеть предполагаются идентичными. В сеть поступает пуассоновский входной поток с параметром $\Lambda(t)$, в общем случае зависящем от времени. Заявка, поступившая в сеть, направляется на обслуживание в узел j с вероятностью r_{0j} , $\sum_{j=1}^m r_{0j} = r \leq 1$. В узле j заявка обслуживается с интенсивностью $\mu_j(n_j)$, где n_j - количество заявок в узле j . $X_j(t)$ - количество заявок в узле j в момент времени t , состояние системы в момент t характеризуется вектором $X(t) = (X_1(t), X_2(t), \dots, X_m(t))$. Для процесса $X(t)$ существует единственное стационарное распределение

$$C_i = 1 + \sum_{n=1}^{\infty} \frac{\lambda_i^n}{\prod_{y=1}^n \mu_i(y)}, i = 1, 2, \dots, m,$$

когда система уравнений трафика:

$$\lambda_i = \Lambda * r_{0i} + \sum_{j=1}^m \lambda_j r_{ji}, i = 1, 2, \dots, m$$

имеет единственное решение [17].

2.2. Модель ненадёжной сети Джексона с динамической маршрутизацией

Теперь предположим, что узлы в сети ненадёжны и могут отказывать и восстанавливаться. Отказы могут быть как индивидуальными, так и групповыми (как в моделях в работах [17, 20, 21]). Пусть $M_0 = \{0, 1, 2, \dots, m\}$ - множество узлов, где 0 - "внешний узел"(вход и выход из сети). Обозначим через $D \subset M$ подмножество отказавших узлов, через $I \subset M \setminus D$ - подмножество работающих узлов, узлы из подмножества I могут отказывать с некоторой интенсивностью $\alpha_{D \cup I}^D(n_i(t))$, узлы $H \subset D$ из подмножества отказавших узлов восстанавливаются с интенсивностью $\beta_{D \setminus H}^D(n_i(t))$.

Пусть (r_{ij}) - исходная матрица маршрутизации. Дополнительно введем матрицу смежности графа сети (s_{ij}) :

$$s_{ij} = \begin{cases} 1, & \text{если } r_{ij} \neq 0, \\ 0, & \text{если } r_{ij} = 0. \end{cases}$$

Рассмотрим множество путей на графе сети. Для этого найдём матрицы: $(s_{ij})^2, (s_{ij})^3, \dots, (s_{ij})^m, m < \infty, (s_{ij})^1 = (s_{ij})$. Известно, что матрица $(s_{ij})^m$ обладает следующим свойством: элемент в строке i и столбце j равен числу путей из узла i в узел j длины m (включающих $(m - 1)$ промежуточных узлов) [23].

Примем следующую схему маршрутизации в сети для узлов из подмножества D (назовем её условно, по аналогии с работой [17], "динамической маршрутизацией без блокировки"), для узлов из подмножества D переходы возможны только в работоспособные узлы $M_0 \setminus D$:

$$r_{ij}^D = \begin{cases} 0, \text{ если } j \in D, i \neq j, \\ r_{ij} + r_{ik}/s_{ik}^p, \text{ если } j \notin D, k \in D \\ \text{и } \exists i \rightarrow j \rightarrow i' \rightarrow j' \rightarrow \dots \rightarrow i'' \rightarrow k : \underbrace{s_{ij}^1 * s_{ji'}^1 * s_{i'j'}^1 * \dots * s_{i''k}^1}_{p+1} \neq 0, \\ \text{где } p = \min\{2, 3, \dots, m : s_{ik_{k \in D}}^p \neq 0\}, \\ r_{ii} + \sum_{\substack{k \in D \\ s_{ik}^p = 0 \forall 1 < p \leq m}} r_{ik}, \text{ если } i \in M_0 \setminus D, i = j, \end{cases}$$

где s_{ik}^p - элемент матрицы $(s_{ij})^p$.

Для входного потока примем такой же принцип маршрутизации:

$$\Lambda r_{0j}^D = \begin{cases} \Lambda r_{0j}, j \in M \setminus D, \\ \Lambda(r_{0j} + r_{0k}/s_{0k}^p * \underbrace{(s_{0j}^1 * s_{ji'}^1 * s_{i'j'}^1 * \dots * s_{i''k}^1)}_{p+1}), \text{ если } j \notin D, k \in D \\ 0, \text{ в остальных случаях.} \end{cases}$$

В дальнейшем через $R^D = (r_{ij}^D)$ будем обозначать модифицированную матрицу маршрутизации. Интенсивности отказов и восстановлений зависят от состояния узлов и не зависят от загрузки сети и определяются следующим образом [17]:

$$\alpha(D, I) = \frac{\psi(D \cup I)}{\psi(D)},$$

$$\beta(D, H) = \frac{\phi(D)}{\phi(D \setminus H)},$$

где ψ, ϕ - положительные функции, определённые на всех подмножествах множества узлов и принимающие конечные значения на конечных подмножествах ($\psi(\emptyset) := 1, \phi(\emptyset) := 1$).

Для узлов сети рассмотрим более общую модель, чем в [20]. Предположим, что каждый узел сети - система массового обслуживания типа $M/G/1$. Динамика сети будет описываться случайным процессом с непрерывным временем $X(t)$, принимающим значения в пространстве состояний \mathbb{E} :

$$\tilde{\mathbf{n}} = ((n_1, z_1), (n_3, z_2), \dots, (n_m, z_m), D) \in \{\mathbb{Z}_+ \times \{R_+ \cup 0\}\}^m \times |D| = \mathbb{E},$$

где n_i - число заявок в узле i , z_i - время, прошедшее от начала обслуживания заявки в узле i , $|D|$ - число элементов в множестве D . Интенсивности обслуживания $\mu_i(n_i, z_i)$ в узлах зависят от числа заявок в системе $n_i(t)$ и от времени $z_i(t)$, прошедшего с начала обслуживания текущей заявки в момент времени t . Рассмотрим условную вероятность отсутствия какого-либо скачка процесса в интервале времени $[t, t + \Delta t)$ ($= \{\text{в сеть не поступит новая заявка}\} \cup \{\text{не завершится обслуживание текущей заявки ни в одном из узлов}\}$ при условии, что текущее значение процесса $X(t)$):

$$\exp\left(-\int_t^{t+\Delta t} \left(\Lambda(s) + \sum_{i=1}^m \mu_i(n_i(t), z_i(t) + s)\right) ds\right),$$

которая, если Δt достаточно мало, определяется как [24, Гл. 2-4]

$$1 - \int_t^{t+\Delta t} \left(\Lambda(s) + \sum_{i=1}^m \mu_i(n_i(t), z_i(t) + s)\right) ds + O(\Delta t)^2,$$

где при $\Delta t \rightarrow 0$ члены с $O((\Delta t)^2)$ пренебрежимо малы по сравнению с членами без Δt или содержащими Δt в первой степени. Аналогично определяются вероятности одного скачка (завершения обслуживания в одном из узлов или прихода одного требования в сеть):

$$\mu_j(n_j(t), z_j(t))\Delta t \left(1 - \int_t^{t+\Delta t} \left(\Lambda(s) + \sum_{i \neq j}^m \mu_i(n_i(t), z_i(t) + s)\right) ds + O(\Delta t)^2\right),$$

$$\Lambda(t)\Delta t \left(1 - \int_t^{t+\Delta t} \left(\sum_{i=1}^m \mu_i(n_i(t), z_i(t) + s)\right) ds + O(\Delta t)^2\right).$$

В сети возможны следующие переходы

$$T_{ij} \tilde{\mathbf{n}} := (D, n_1, \dots, n_i - 1, \dots, n_j + 1, \dots, n_m),$$

$$\begin{aligned}
T_{0j}\tilde{\mathbf{n}} &:= (D, n_1, \dots, n_j + 1, \dots, n_m), \\
T_{i0}\tilde{\mathbf{n}} &:= (D, n_1, \dots, n_i - 1, \dots, n_m), \\
T_H\tilde{\mathbf{n}} &:= (D \setminus H, n_1, \dots, n_m), \\
T^I\tilde{\mathbf{n}} &:= (D \cup I, n_1, \dots, n_m).
\end{aligned}$$

Определение 1 Будем называть ненадёжной сетью массового обслуживания (СеМО) марковский процесс $\mathbf{X} = (X(t), t \geq 0)$, определяемый следующим инфинитезимальным генератором:

$$\begin{aligned}
\tilde{Q}f(\tilde{\mathbf{n}}) = & \sum_{j=1}^m [f(T_{0j}\tilde{\mathbf{n}}) - f(\tilde{\mathbf{n}})]\Lambda(t)r_{0j}^D \\
& + \sum_{i=1}^m \sum_{j=1}^m [f(T_{ij}\tilde{\mathbf{n}}) - f(\tilde{\mathbf{n}})]\mu_i(n_i(t), z_i(t))r_{ij}^D \\
& + \sum_{I \subset M} [f(T^I\tilde{\mathbf{n}}) - f(\tilde{\mathbf{n}})]\alpha(D, I) + \sum_{H \subset M} [f(T^I\tilde{\mathbf{n}}) - f(\tilde{\mathbf{n}})]\beta(D, H) \\
& + \sum_{j=1}^m [f(T_{j0}\tilde{\mathbf{n}}) - f(\tilde{\mathbf{n}})]\mu_j(n_j(t), z_j(t))r_{j0}^D.
\end{aligned}$$

3. Основные результаты

По аналогии с классической сетью Джексона и для случая с блокировками [17] можно показать, что для сети с динамической маршрутизацией существует стационарное распределение.

Теорема 1 Пусть для определённой выше ненадёжной СеМО выполнены условия

$$1) \inf_{n_j, t} \mu_j(n_j, z_j) > 0 \quad \forall j,$$

2) времена обслуживания и поступления заявок независимые случайные величины,

3) матрица R^D неразложима,

тогда для СеМО существует стационарное распределение, задаваемое формулами

$$\pi(\tilde{\mathbf{n}}) = \pi(D, n_1, n_2, \dots, n_m) = \frac{1}{C} \frac{\psi(D)}{\phi(D)} \prod_{i=1}^m \frac{1}{C_i} \frac{\lambda_i^{n_i}}{\prod_{k=1}^{n_i} \mu_i(k)}$$

где

$$C_i = \sum_{n=0}^{\infty} \frac{\lambda_i^n}{\prod_{y=1}^n \mu_i(y)}, \quad \lambda_i = \sum_{j=0}^m \Lambda * r_{ji}.$$

Сформулируем основные результаты для оценки скорости сходимости для ненадёжной СеМО через оценку спектрального зазора. Приведем предварительные сведения о спектральном зазоре: пусть $\mathbf{X} = (X_t, t \geq 0)$ марковский процесс, $Q = [q(\mathbf{e}, \mathbf{e}')]_{\mathbf{e}, \mathbf{e}' \in \mathbb{E}}$ - матрица интенсивностей перехода, π - стационарное распределение процесса и инфинитезимальный генератор определяется как

$$\mathbf{Q}f(\mathbf{e}) = \sum_{\mathbf{e}' \in \mathbb{E}} (f(\mathbf{e}') - f(\mathbf{e}))q(\mathbf{e}, \mathbf{e}').$$

Скалярное произведение на $L_2(\mathbb{E}, \pi)$: $\langle f, g \rangle_{\pi} = \sum_{\mathbf{e} \in \mathbb{E}} f(\mathbf{e})g(\mathbf{e})\pi(\mathbf{e})$. Спектральный зазор для \mathbf{X} определяется следующим образом:

$$\text{Gap}(\mathbf{Q}) = \inf\{-\langle f, \mathbf{Q}f \rangle_{\pi} : \|f\|_2 = 1, \langle f, \mathbf{1} \rangle_{\pi} = 0\}.$$

Основные результаты сформулированы в следующих теоремах:

Теорема 2 Пусть \mathbf{X} - марковский процесс с инфинитезимальным генератором \mathbf{Q} , предположим, что \mathbf{Q} - ограничен, минимальная интенсивность обслуживания $\inf_{n_j, t} \mu_j(n_j, z_j) > 0$ и матрица маршрутизации (r_{ij}^D) неразложима, тогда $\text{Gap}(\mathbf{Q}) > 0$, если выполнено следующее условие: при всех $i = 1, \dots, t$, для процесса рождения-гибели, соответствующего работе i -го узла сети с параметрами λ_i and $\mu_i(n_i, z_i)$ выполнено условие $\text{Gap}_i(\mathbf{Q}_i) > 0$.

Теорема 3 Пусть \mathbf{X} марковский процесс с ограниченным инфинитезимальным генератором \mathbf{Q} , минимальной интенсивностью обслуживания $\inf_{n_j, t} \mu_j(n_j, z_j) > 0$ и неразложимой матрицей маршрутизации (r_{ij}^D) , тогда $\text{Gap}(\mathbf{Q}) > 0$ т. и т.т., когда для всех $i = 1, \dots, t$, распределение $\pi = (\pi_i), i \geq 0$ имеет легкие хвосты, т.е. выполнено условие $\inf_k \frac{\pi_i(k)}{\sum_{j>k} \pi_i(j)} > 0$.

Теорема 4 (следствие результата [22]) Пусть \mathbf{X} ненадёжная СеМО, определённая выше, с инфинитезимальным генератором \mathbf{Q} и соответствующей матрицей переходных вероятностей P_t . Предположим, что матрица маршрутизации (r_{ij}^D) неразложима и $(r_{ij}^D)^k > 0$ для $k \geq 1$. Если распределение π_i имеет легкие хвосты для всех $i = 1, \dots, t$, тогда следующие условия эквивалентны

– для всех $f \in L_2(\mathbb{E}, \pi)$

$$\|P_t f - \pi(f)\|_2 \leq e^{-\text{Gap}(\mathbf{Q})t} \|f - \pi(f)\|_2, t > 0,$$

– для любого $\mathbf{e} \in \mathbb{E}$ существует константа $C(\mathbf{e}) > 0$ такая, что

$$\|\delta_{\mathbf{e}} - \pi(f)\|_{TV} \leq C(\mathbf{e})e^{-\text{Gap}(\mathbf{Q})t}, t > 0.$$

Благодарности

Работа выполнена при поддержке РФФИ (гранты № 14-07-31245 и № 15-08-08677).

Литература

1. *Laszlo Lakatos, Laszlo Szeidl, Miklos Telek.* Introduction to Queueing Systems with Telecommunication Applications. Springer Science & Business Media, Dec 15, 2012 - Mathematics - 388 pages
2. *A. Ghosal, S. C. Gujaria, Raja Ghosal.* Network Queueing Systems: With Industrial Applications. South Asian Publishers, Jan 1, 2004 - Computer network protocols - 138 pages
3. *John Daigle.* Queueing Theory with Applications to Packet Telecommunication. Springer Science & Business Media, Jan 16, 2006 - Technology & Engineering - 316 pages
4. *Thomasian Alexander.* Analysis of Fork/Join and Related Queueing Systems. ACM Comput. Surv. 47, 2, Article 17 (August 2014), 71 pages. DOI=<http://dx.doi.org/10.1145/2628913>
5. *Lakshmi C., Sivakumar A. I.* Application of queueing theory in health care: A literature review // Operations Research for Health Care Volume 2, Issues 1–2, March–June 2013, Pages 25–39
6. *Sterbenz, J.P.G., Cetinkaya, E.K., Hameed, M.A. et al.* Evaluation of network resilience, survivability, and disruption tolerance: analysis, topology generation, simulation, and experimentation // Telecommun Syst (2013) 52: 705. doi:10.1007/s11235-011-9573-6
7. *I-Hong Hou, P. R. Kumar.* Queueing systems with hard delay constraints: a framework for real-time communication over unreliable wireless channels // Queueing Systems, 2012, Volume 71, Issue 1, pp 151-177. doi:10.1007/s11134-012-9293-y.
8. *Madhu Jain, G.C. Sharma, Richa Sharma.* Unreliable server M/G/1 queue with multi-optional services and multi-optional vacations // International Journal of Mathematics in Operational Research, Volume 5, Issue 2, DOI: 10.1504/IJMOR.2013.052458
9. *Omar M. Ba-Rukab, Lotfi Tadj, Jau-Chuan Ke.* Binomial schedule for an M/G/1 queueing system with an unreliable server // International Journal of Modelling in Operations Management. Volume 3, Issue 3-4, DOI: 10.1504/IJMOM.2013.058326
10. *Lotfi Tadj, Gautam Choudhury, Kamel Rekab.* A two-phase quorum queueing system with Bernoulli vacation schedule, setup, and N-policy for an unreliable server with delaying repair // International Journal of Services and Operations Management, 2012, Volume 12, Issue 2, DOI: 10.1504/IJSOM.2012.047103
11. *S. Bama, M. I. Aftab Begum, P. Fijy Jose.* Unreliable Mx/G/1 Queueing System with Two Types of Repair // International Journal of Innovative Research and Development, Volume 4, Issue 10, 2015

12. *Valentina Klimenok , Vladimir Vishnevsky.* Unreliable Queueing System with Cold Redundancy. Chapter in Computer Networks (Editors Piotr Gaj, Andrzej Kwiecień, Piotr Stera). Volume 522 of the series Communications in Computer and Information Science, pp 336-346, 2015, doi: 10.1007/978-3-319-19419-6_32
13. *Hong Chen, David D. Yao.* Fundamentals of Queueing Networks. Performance, Asymptotics, and Optimization. - Springer, Book on Stochastic Modelling and Applied Probability, Volume 46, 2001, DOI 10.1007/978-1-4757-5301-1
14. *Vvedenskaya, N.D.* Configuration of overloaded servers with dynamic routing // Probl Inf Transm (2011) 47: 289. doi:10.1134/S0032946011030070.
15. *Sukhov Yu.M., Vvedenskaya, N.D.* Fast Jackson Networks with Dynamic Routing // Problems of Information Transmission (2002) 38: 136. doi:10.1023/A:1020010710507.
16. *Kelly F.P., Laws C.N.* Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling // Queueing Systems, 13 (1993), 47-86.
17. *Sauer C. , Daduna H.* Availability Formulas and Performance Measures for Separable Degradable Networks // Economic Quality Control, Vol 18 (2003), No. 2, 165 – 194
18. *Conte Marco.* Dynamic Routing in Broadband Networks // Springer Science & Business Media, 2012 - Computers.
19. *F. Yavuz and J. Zhao and O. Yağan and V. Gligor.* Toward k-Connectivity of the Random Graph Induced by a Pairwise Key Predistribution Scheme With Unreliable Links // IEEE Transactions on Information Theory, 2015, vol. 61, issue 11, pp. 6251-6271, doi: 10.1109/TIT.2015.2471295.
20. *Lorek P., Szekli R.* Computable bounds on the spectral gap for unreliable Jackson networks. // Adv. in Appl. Probab, Volume 47, Number 2 (2015), 402-424.
21. *Lorek P.* The exact asymptotic for the stationary distribution of some unreliable systems. // arXiv:1102.4707 [math.PR].
22. *Chen, M.F.* Eigenvalues, Inequalities, and Ergodic Theory. Springer, 2005.
23. *Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein.* Introduction to Algorithms, Third Edition (3rd ed.). The MIT Press, 2009
24. *Saaty T.L.* Elements of Queueing theory with applications. NY, Dover Publications, 1983.

УДК 519.245

Оценивание эффективной пропускной способности узла высокоответственной сети

К. А. Калинина

*Институт прикладных математических исследований,
Карельский научный центр РАН,
ул. Пушкинская, д.11, Петрозаводск, Россия*

Аннотация. В статье рассматривается регенеративный метод оценивания эффективной пропускной способности (ЭПС) узла тандемной сети и обсуждается переоценивание, которое наблюдается в результатах имитационного моделирования. Под переоцениванием понимается избыточная пропускная способность, которая удовлетворяет некоторому заданному требованию качества обслуживания (QoS). Особое внимание уделяется возможной причине появления этого переоценивания — случайному суммированию нагрузки на циклах. Полученные результаты моделирования позволяют сделать определенные выводы о возможности использования регенеративного метода и непосредственно случайного суммирования для оценивания ЭПС узла высокоответственной сети.

Ключевые слова: эффективная пропускная способность, регенеративное оценивание, QoS, высокоответственная сеть.

1. Введение

В статье рассматривается проблема оценивания эффективной пропускной способности (ЭПС) узла высокоответственной сети регенеративным методом. Ранее этот вопрос частично обсуждался в [2–5]. Напомним, что под ЭПС понимается такое значение пропускной способности C обслуживающего прибора, для которого выполняется следующее требование качества обслуживания: вероятность превышения стационарным процессом нагрузки W некоторого (заданного) высокого уровня b не превышает (заданного) малого значения

$$P_b := \mathbb{P}(W > b) \leq \Gamma.$$

Коммуникационные сети и системы, для которых нарушение введенного в рассмотрение требования (1) является крайне нежелательным (и даже критичным), будем называть высокоответственными. Предыдущие исследования возможных способов оценивания ЭПС показали, что самой предпочтительной из изученных является регенеративная оценка. Подробнее ознакомиться со сравнительным анализом разных методов оценивания ЭПС на основе имитационного моделирования можно, например, в [2, 3, 5]. В этой работе мы остановимся на обсуждении ранее установленного переоценивания Δ (избыточной

пропускной способности) в случае использования этого регенеративного метода в сетях разных типов [2,4] ($10\% \leq \Delta \leq 17\%$). В нашем случае переоценивание означает, что регенеративная оценка гарантирует выполнение требования (1) с небольшим запасом. В статье обсуждается основная причина появления этого переоценивания.

2. Случайное суммирование в оценивании ЭПС

Несмотря на то, что установленное ранее переоценивание является допустимым в всокоответственных системах и сетях, не следует оставлять без внимания причины его появления. Предыдущие исследования, посвященные сравнению разных способов оценивания ЭПС [3–5], дают основания предположить что возможной и наиболее вероятной причиной появления избыточной пропускной способности является случайное суммирование данных, принадлежащий одному циклу регенерации. Для проверки этого предположения построим оценку ЭПС, основанную на искусственных случайных циклах (вместо регенеративных циклов, которые получаются естественным путем). Чтобы изучить свойства случайного суммирования, будем менять распределение длин искусственных циклов.

Мы моделируем тандемную сеть из двух узлов (выходной процесс первого узла создает регенеративный входной процесс во второй узел, для которого и необходимо рассчитать ЭПС). В таблице 1 представлены результаты моделирования для длин искусственных циклов с распределением Вейбулла с функцией $F(x) = 1 - e^{-0.3x^2}$. Таблица 2 содержит результаты для случая, когда длины циклов имеют распределение Парето с функцией $F(x) = 1 - (\frac{1}{x})^{1.5}$ (для $x \geq 1$).

Таблица 1

Искусственные циклы, Weibull β

Γ	\hat{C}_{rnd}	$\hat{\Gamma}_{rnd}$	Δ_{rnd}	$\hat{\Gamma}_{\Delta_{rnd}}$
10^{-4}	5.70	$1.39 \cdot 10^{-5}$	0.14	$2.58 \cdot 10^{-4}$
10^{-5}	6.16	$4.78 \cdot 10^{-7}$	0.16	$4.22 \cdot 10^{-5}$
10^{-6}	6.42	$2.08 \cdot 10^{-8}$	0.18	$1.05 \cdot 10^{-6}$

Результаты моделирования показали, что переоценивание в случае искусственных циклов сильно зависит от типа распределения длин циклов и в целом сравнимо с переоцениванием в случае использования регенерации: $0.14 \leq \Delta_{rnd} \leq 0.18$ (Напомним, что для для регенеративных циклов переоценивание в среднем принимает значения $10\% \leq \Delta \leq 17\%$ [4]). Следует заметить, что в случае с распределением Парето (см. таблицу 2) переоценивание достигает достаточно высоких

Таблица 2

Искусственные циклы, Pareto β

Γ	\hat{C}_{rnd}	$\hat{\Gamma}_{rnd}$	Δ_{rnd}	$\hat{\Gamma}_{\Delta_{rnd}}$
10^{-4}	5.42	$4.23 \cdot 10^{-5}$	0.20	$6.12 \cdot 10^{-4}$
10^{-5}	6.33	$3.75 \cdot 10^{-7}$	0.24	$8.09 \cdot 10^{-5}$
10^{-6}	5.79	$1.88 \cdot 10^{-8}$	0.25	$7.91 \cdot 10^{-6}$

значений 25%- 28%. Подобные значения хоть и допустимы (с точки зрения использования этого метода в высокоответственных сетях), но ставят под сомнение рациональность использования искусственных циклов. В связи с этим регенеративная оценка все же остается более предпочтительной.

3. Заключение

Результаты моделирования, представленные в данной статье, показали, что наиболее вероятной причиной появления переоценивания (избыточной пропускной способности) является случайное суммирование данных на циклах регенерации. Построенная на основе искусственных случайных циклов оценка ЭПС сильно зависит от распределения количества слагаемых. В некоторых случаях это приводит к нежелательному увеличению переоценивания. Регенеративный метод, в свою очередь, переоценивает ЭПС в относительно разумных (по сравнению с оценкой, построенной на случайном суммировании) пределах, что позволяет считать его более предпочтительным. Оба метода являются особенно актуальными для использования в высокоответственных сетях, где переоценивание ЭПС допустимо, а недооценивание крайне критично. В дальнейших исследованиях мы планируем использовать идею случайного суммирования для возможного улучшения регенеративной оценки ЭПС.

Благодарности

Работа частично поддержана грантами РФФИ № 15-07-02341, № 15-07-02354, № 15-07-02360, а также программой стратегического развития Петрозаводского государственного университета на 2012 – 2016 годы.

Литература

1. *Parker L., Christensen S. M.* MathTensor: a system for doing tensor analysis by computer. — Addison-Wesley, 1994.

2. *Borodina A., Kalinina K., Morozov E.* On the accuracy of the effective bandwidth regenerative estimation, ICUMT, Saint-Petersburg: 2014. P. 652-656.
3. *Kalinina K., Morozov E., Rykov V.* Effective bandwidth estimation in highly reliable regenerative networks // The Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management, Israel, 2016: Proceedings, P. 323-327.
4. *Morozov E., Kalinina K.* On the effective bandwidth estimation in communication network // 29th European Conference on Modelling and Simulation ECMS 2015: Proceedings, P.423-429.
5. *Калинина К.* Об эффективной пропускной способности узлов коммуникационной сети, Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем: материалы Всероссийской конференции с международным участием.— Москва: РУДН, 2015, 24-26.

UDC 519.245

Effective bandwidth estimation in highly critical systems

К. А. Kalinina

*Institute of Applied Mathematical Research
Karelian Research Center of Russian Academy of Sciences
Pushkinskaya str. 11, Petrozavodsk, Russia*

In this paper, we study the accuracy of regenerative estimator of the effective bandwidth. It is a required server capacity to guarantee a given QoS requirement. By simulation we show that this method overestimates predefined QoS requirement. We study the possible reason of this overestimation (random sums of the workload on the cycles) and discuss the opportunity to apply regenerative estimator for effective bandwidth estimation in highly critical systems.

Keywords: effective bandwidth, QoS, highly critical systems, regeneration.

UDC 517.937, 517.928.2, 519.711.2

Neural networks with an infinite number of cells analysis

S. K. Kanzitdinov, S. A. Vasilyev

*Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

Abstract. A method of analysis the dynamics of complex systems using neural networks with an infinite number of cells was investigated. For the Cauchy problem for systems of differential equations of countable order, which describes the neural network with infinite number of cells, considered the question of the existence and uniqueness of its solution.

Keywords: neural network; forecasting of the dynamics of complex systems.

1. Introduction

The recent research of the dynamics of complex systems using neural networks with an infinite number of cells in faced with the problem of the solutions analysis of certain infinite systems of ordinary differential equations to a time-independent solution. A model for a large network of "neurons" with a graded response (or sigmoid input-output relation) was studied [4]. The idea was used in biological systems was given added credence by the continued presence of such properties for more nearly biological "neurons". In the paper [1] was given existence and uniqueness results for the equations describing the dynamics of some neural networks for which there were infinitely many cells. Such system was considered and neural nets which were modelled were described by the infinite system of ordinary differential equations.

Bruce D. Calvert and Armen H. Zemanian [2] investigated a nonlinear infinite resistive network, an operating point could be determined by approximating the network by finite networks obtained by shorting together various infinite sets of nodes, and then taking a limit of the nodal potential functions of the finite networks. By taking a completion of the node set of the infinite network under a metric given by the resistances, limit points were obtained that represent generalized ends, which it be called "terminals," of the infinite network. These terminals could be shorted together to obtain a generalized kind of node, a special case of a 1-node. An operating point will involve Kirchhoff's current law holding at 1-nodes, and so the flow of current into these terminals was studied. They gave existence and bounds for an operating point that also had a nodal potential function, which was continuous at the 1-nodes. The existence was derived from the said approximations.

Haiying Huang, Qiaosheng Du and Xibing Kang [5] studied a class of neutral high-order stochastic Hopfield neural networks with Markovian

jump parameters and mixed time delays. The jumping parameters was modeled as a continuous-time finite-state Markov chain. The existence of equilibrium point for the addressed neural networks was studied. By utilizing the Lyapunov stability theory, stochastic analysis theory and linear matrix inequality (LMI) technique, new delay-dependent stability criteria were presented in terms of linear matrix inequalities to guarantee the neural networks to be globally exponentially stable in the mean square.

In paper [15] Xiao Liang, Linshan Wang, Yangfan Wang and Ruili Wang focused on the long time behavior of the mild solution to delayed reaction-diffusion Hopfield neural networks (DRDHNNs) driven by infinite dimensional Wiener processes. They analyzed the existence, uniqueness, and stability of this system under the local Lipschitz function by constructing an appropriate Lyapunov-Krasovskii function and utilizing the semigroup theory. Some easy-to-test criteria affecting the well-posedness and stability of the networks, such as infinite dimensional noise and diffusion effect, were obtained. The criteria could be used as theoretic guidance to stabilize DRDHNNs in practical applications when infinite dimensional noise was taken into consideration. Considering the fact that the standard Brownian motion is a special case of infinite dimensional Wiener process, they undertake an analysis of the local Lipschitz condition, which had a wider range than the global Lipschitz condition.

In 1984 Hopfield investigated a neural network which was described using system of ordinary differential equations [4]

$$C_i \frac{du_i}{dt} = \sum_{j=1}^N T_i^j g(u_j(t)) + d_i - \frac{u_i(t)}{R_i}, \quad (1)$$

$$i, j = 1, \dots, N, t \geq 0$$

or

$$\frac{du_i}{dt} = f_i(u, g, t) + r_i, \quad i, j = 1, \dots, N,$$

$$f_i(u, g, t) = \sum_{j=1}^N T_i^j g(u_j)/C_i, \quad r_i = (d_i - u_i(t)/R_i)/C_i,$$

where $u_i(t)$ are the voltage changes on the neuron, which determines the state of the system, $C_i > 0$, $R_i > 0$, T_{ij} , d_i are sets of real numbers, and a function $g \in [-1, 1]$, ($s \in R$) increasing function. In this work, we study the following system of singularly perturbed differential equations, which is a generalization of system (1):

$$\begin{cases} \frac{du_i}{dt} = f_i(u, g, t) + m_i, & i = 1, \dots, N, \\ \mu \frac{du_i}{dt} = F_i(u, g, t) + M_i, & i = N + 1, \dots, \end{cases} \quad (2)$$

$$\begin{cases} f_i(u, g, t) = \sum_{j=1}^{\infty} T_i^j g(u_j)/C_i, \\ m_i = (d_i - u_i(t)/R_i)/C_i, \\ i = 1, \dots, N, \end{cases}$$

$$\begin{cases} F_i(u, g, t) = \sum_{j=1}^{\infty} T_i^j g(u_j)/C_i, \\ M_i = (d_i - u_i(t)/R_i)/C_i, \\ i = 1, 2, \dots, \end{cases}$$

where $u_i(t)$, $i = 1, 2, \dots$ is a functional sequence, $C_i > 0$, $R_i > 0$, T_{ij} , $d_i \subset l_1$ are numerical sequence, and $\mu > 0$ is a small parameter.

For system (2) we can formulate the following Cauchy problem:

$$\begin{cases} \frac{du_i}{dt} = f_i(u, g, t) + m_i, \quad i = 1, \dots, N, \\ \mu \frac{du_i}{dt} = F_i(u, g, t) + M_i, \quad i = N + 1, \dots, \\ u_i(0) = \bar{u}_i^0, \quad i = 1, 2, \dots, \end{cases} \quad (3)$$

where $\bar{u}_i^0 \subset l_1$ numerical sequence.

Cauchy problems for the systems of ordinary differential equations of infinite order was investigated A.N.Tihonov [10], K.P.Persidsky [9], O.A.Zhautykov [13], [14], Ju.Korobeinik [6] other researchers.

It was studied the singular perturbed systems of ordinary differential equations by A.N. Tihonov [11], A.B.Vasil'eva [12], S.A. Lomov [8] other researchers.

In this paper the Cauchy problem (3) is considered the existence and uniqueness of its solution, an algorithm for constructing asymptotic solutions using approximate methods of solutions of differential equations with a small parameter at the highest derivative and analyzed the possibility of applying it solutions to predict the dynamics of complex systems in conditions of uncertainty.

For the Cauchy problem for systems of differential equations of countable order, which describes the neural network with infinite number of cells, considered the question of the existence and uniqueness of its solution, an algorithm for constructing asymptotic solutions using approximate methods of solutions of differential equations with a small parameter at the highest derivative and analyzed the possibility of applying it solutions to predict the dynamics of complex systems in conditions of uncertainty.

In this paper we apply methods from [12] for the singular perturbed systems of ordinary differential equations of infinite order of Tikhonov-type.

2. Tikhonov-type Cauchy problems for systems of ordinary differential equations of infinite order with a small parameter

Let's consider Tikhonov-type Cauchy problems for systems of ordinary differential equations of infinite order with a small parameter μ and initial conditions (3):

$$\begin{cases} \dot{x} = f(x(t), y(t), t) + m, \\ \mu \dot{y} = F(x(t), y(t), t) + M; \\ x(t_0) = u_x^0, y(t_0) = u_y^0, \end{cases} \quad (4)$$

$$\begin{aligned} x &= (u_1, \dots, u_N)^T, y = (u_{N+1}, u_{N+2}, \dots)^T, \\ f &= (f_1, \dots, f_N)^T, F = (F_{N+1}, F_{N+2}, \dots)^T, \\ m &= (m_1, \dots, m_N)^T, M = (M_{N+1}, M_{N+2}, \dots)^T, \\ u_x^0 &= (\bar{u}_1^0, \dots, \bar{u}_N^0)^T, u_y^0 = (\bar{u}_{N+1}^0, \bar{u}_{N+2}^0, \dots)^T, \end{aligned}$$

where $x, f, m \in X$, $X \in \mathbf{R}^n$ are n -dimensional functions; $y, F, M \in Y$, $Y \subset l_1$ are infinite-dimensional functions and $t \in [t_0, t_1]$ ($t_0 < t_1 \leq \infty$), $t \in T$, $T \in \mathbf{R}$; $u_x^0 \in X$ and $u_y^0 \in Y$ are given vectors, $\mu > 0$ is a small real parameter; $x(t) = (x_1, \dots, x_N)^T$ and $y(t) = (y_1, y_2, \dots)^T$ are solutions of (4). Given functions $f(x(t), y(t), t) = (f_1, \dots, f_N)^T$ and $F(x(t), y(t), t) = (F_1, F_2, \dots)^T$ are continuous functions for all variables.

Let S is an integral manifold of the system (4) in $X \times Y \times T$. If any point $t^* \in [t_0, t_1]$ $(x(t^*), y(t^*), t^*) \in S$ of trajectory of this system has at least one common point on S this trajectory $(x(t), y(t), t) \in S$ belongs the integral manifold S totally. If we assume in (4) that $\mu = 0$ than we have a degenerate system of the ordinary differential equations and a problem of singular perturbations

$$\begin{cases} \dot{x} = f(x(t), y(t), t) + m, \\ 0 = F(x(t), y(t), t); \\ x(t_0) = u_x^0, \end{cases} \quad (5)$$

where the dimension of this system is less than the dimension of the system (4), since the relations $F(x(t), y(t), t) = 0$ in the system (5) are the algebraic equations (not differential equations). Thus for the system (5) we can use limited number of the initial conditions then for system (4). Most natural for this case we can use the initial conditions $x(t_0) = u_x^0$ for the system (5) and the initial conditions $y(t_0) = u_y^0$ disregard otherwise we get the overdefined system. We can solve the system (5) if the equation

$F(x(t), y(t), t) = 0$ could be solved. If it is possible to solve we can find a finite set or countable set of the roots $y_q(t) = w_q(x(t), t)$ where $q \in \mathbf{N}$.

If the implicit function $F(x(t), y(t), t) = 0$ has not simple structure we must investigate the question about the choice of roots. Hence we can use the roots $y_q(t) = w_q(x(t), t)$ ($q \in \mathbf{N}$) in (5) and solve the degenerate system

$$\begin{cases} \dot{x}_d = f(x_d(t), w_q(x_d(t), t), t) + m; \\ y_d(t_0) = u_x^0. \end{cases} \quad (6)$$

Since it is not assumed that the roots $y_q(t) = w_q(x(t), t)$ satisfy the initial conditions of the Cauchy problem (4) ($y_q(t_0) \neq w_x$, $q \in \mathbf{N}$), the solutions $y(t)$ (4) and $y_q(t)$ do not close to each other at the initial moments of time $t > 0$. Also there is a very interesting question about behaviors of the solutions $x(t)$ of the singular perturbed problem (4) and the solutions $x_d(t)$ of the degenerate problem (6). When $t = 0$ we have $x(t_0) = x_d(t_0)$. Do these solutions close to each other when $t \in (t_0, t_1]$? The answer to this question depends on using roots $y_q(t) = w_q(x(t), t)$ and the initial conditions which we apply for the systems (4) and (5).

3. Local existence theorem for Cauchy problems for systems of ordinary differential equations of infinite order

Let Tikhonov-type Cauchy problems for systems of ordinary differential equations of infinite order with a small parameter $\mu > 0$ and initial conditions (4) has a form:

$$\begin{cases} \dot{z} = P(z(t, G, \mu), t, \mu) + Q; \\ z(t_0, \mu) = G, \end{cases} \quad (7)$$

$$z = (x_1, x_2, \dots, x_N, y_1, y_2, \dots)^T,$$

$$P(z(t, G, \mu), t, \mu) = (f_1, f_2, \dots, f_N, \mu^{-1}F_1, \mu^{-1}F_2, \dots)^T,$$

$$Q = (m_1, m_2, \dots, m_N, M_{N+1}, M_{N+2}, \dots)^T,$$

$$G = (u_{x_1}^0, u_{x_2}^0, \dots, u_{x_N}^0, u_{y_1}^0, u_{y_2}^0, \dots)^T,$$

where $P(z(t, G, \mu), t, \mu)$ is the infinite-dimensional function; G is the given vector; $t \in [t_0, t_1]$ ($t_0 < t_1 \leq \infty$).

Let $z(t, G, \mu)$ be a continuously differentiable solution of the Cauchy problems (7) then there are $\Phi(t, G, \mu) = \partial z(t, G, \mu) / \partial G$, $\Psi(t, G, \mu) =$

$\partial z(t, G, \mu)/\partial \mu$ where $\Phi(t, G, \mu)$ and $\Psi(t, G, \mu)$ satisfy of the system of ordinary differential equations in variations:

$$\dot{z} = P(z(t, G, \mu), t, \mu),$$

$$\begin{cases} \dot{\Phi}(t, G, \mu) = J_z(t, G, \mu)\Phi(t, G, \mu), \\ \dot{\Psi}(t, G, \mu) = J_z(t, G, \mu)\Psi(t, G, \mu) + \Lambda_\mu(t, G, \mu); \\ z(t_0, G, \mu) = G, \Phi(t_0, G, \mu) = I, \Psi(t_0, G, \mu) = 0, t_0 \in T, \end{cases} \quad (8)$$

where $J_z(t, G, \mu) = (\partial P_i/\partial z_j)_{i,j=1}^\infty$ is Jacobi matrix, I is an identity operator and $\Lambda_\mu(t, G, \mu) = (\partial P_i/\partial \mu)_{i=1}^\infty$ is a vector.

Theorem 1 (local existence theorem). *Let $P(z(t, G, \mu), t, \mu)$, $J_z(t, G, \mu)$, $\Lambda_\mu(t, G, \mu)$ be continuous and meet Gelder's local condition with $z \in U_\epsilon(G)$ then the system (8) has only one solution, which meet the conditions $z(t_0, G, \mu) = G$, $z(t, G, \mu) \in U_\epsilon(G)$. Thus $z(t, G, \mu)$ continuously differentiable with respect to the initial condition, and its derivative meet the equation (8).*

Proof. This statement is following from [3] (theorem 3.4.4) when the unlimited operator be $A = 0$. *End proof.*

The behavior of the solution $z(t, G, \mu)$ (7) and the nonnegative condition for the off-diagonal elements of the matrix $J_z(t, G, \mu)$ is demonstrated by the following theorem.

Theorem 2. *Let the solution z of (7) be $z(t, G, \mu) \in l_1$ for any $t \geq 0$, $G \in l_1$ and μ . The following claims are equal: (i) the off-diagonal elements $J_z(t, G, \mu)$ are non-negative for any G ; (ii) for any G and any vector $h \in l_1$, $h \geq 0$, $z(t, G + h, \mu) \geq z(t, G, \mu)$.*

Proof. Let us examine a convex set Z , and $z(t, G, \mu) \in Z$ for any $G \in Z$, derivative $\Phi(t, G, \mu)$ of function $z(t, G, \mu)$ can be specify by simultaneous equations (8). In that case the following formula is fair for any $G^0, G^1 \in Z$:

$$z(t, G^1, \mu) - z(t, G^0, \mu) = \int_0^1 \Phi(t, \gamma(s), \mu)(G^1 - G^0)ds \quad (9)$$

where $\gamma(s) = (1 - s)G^0 + sG^1, 0 \leq s \leq 1$.

In fact the function $z(t, G, \mu)$ transfer the segment $\gamma(s)$ into the curve $z(t, \gamma(s), \mu)$ in (9). The following formula is fair because of the continuous differentiability of function $z(t, G, \mu)$

$$z(t, \gamma(\tau), \mu) = z(t, G^0, \mu) + \int_0^\tau \frac{\partial z(t, \gamma(s), \mu)}{\partial s} ds.$$

By the formula of complex derivative

$$\frac{\partial z(t, \gamma(s), \mu)}{\partial s} = \frac{\partial z}{\partial G}(\gamma(s))\gamma'(s)$$

Recalling that $\partial z/\partial G = \Phi$ and $\gamma'(s) = G^1 - G^0$, with $\tau = 1$ we get (9). Let us suppose that statement (i) is fair. So because of (9)

$$z(t, G + h, \mu) - z(t, G, \mu) = \int_0^1 \Phi(t, \gamma(s), \mu) h ds$$

where $\gamma(s) = G + sh, 0 \leq s \leq 1$. Because of non-negativeness of function $J_z(t, G, \mu)$ outside of diagonal from (9) we get $\Phi(t, \gamma(s), \mu) \geq 0$, so $\Phi(t, \gamma(s), \mu)h \geq 0$ whence we get statement (ii).

Let us suppose that (ii) is fair. Under the conditions of Theorem 1 P, J_z with $z \in U_\epsilon(G)$ be continuous and meet Gelder's local condition. Let Gelder's local condition be $\|P\| < M_0, \|J\| < M_1$, and there are numbers $\delta > 0, \delta = \min(\epsilon/M_0, 1/M_1)$. Let $z(t, G, \mu) = G + z^*(t, G, \mu)$ be a solution of (9), where $z^*(t, G, \mu)$ is a fixed point of Picard's mapping $(\prod \theta)(t) = \int_{t_0}^t P(G + \theta(\tau)) d\tau$ under conditions $t \in [t_0 - \delta_1, t_0 + \delta_1], \delta_1 < \delta$. Mapping \prod is contraction with coefficient $\lambda = \delta_1 M_1 < 1$. Consider the approximation to solution $\tilde{z}(t, G, \mu) = G + \tilde{z}^*(t, G, \mu) = G + (t - t_0) P(z(t, G, \mu), t, \mu)$. We can see that

$$\begin{aligned} & \|\tilde{z}(t, G, \mu) - z(t, G, \mu)\| = \\ & = \|\tilde{z}^*(t, G, \mu) - z^*(t, G, \mu)\| \leq \\ & \leq \frac{1}{1 - \lambda} \|\prod \tilde{z}(t, G, \mu) - \tilde{z}(t, G, \mu)\|, \\ & \quad \prod \tilde{z}(t, G, \mu) - \tilde{z}(t, G, \mu) = \\ & = \int_{t_0}^t P(G + (\tau - t_0)P) d\tau - \int_{t_0}^t P d\tau = \\ & = \int_{t_0}^t (P(G + (\tau - t_0)P) - P) d\tau = D. \end{aligned}$$

Because of the derivative of the function P is limited and P meet Gelder's local condition with the constant M_1 , where $\|P(G + (\tau - t_0)P(G)) - P(G)\| \leq M_1 \|(\tau - t_0)P(G)\| \leq M_0 M_1 |\tau - t_0|$, so $\|D\| \leq M_0 M_1 (t - t_0)^2 / 2(1 - \lambda)$, or $\|\tilde{z}(t, G, \mu) - z(t, G, \mu)\| \leq M_0 M_1 (t - t_0)^2 / 2(1 - \lambda)$. Using this estimation and for all small $\zeta > 0$ we have that

$$0 \leq z(t, G + \zeta e_j, \mu) - z(t, G, \mu) = \zeta e_j + (t - t_0)[P(G + \zeta e_j) - P(G)] + \gamma(G, t),$$

where $\|\gamma(G, t)\| \leq M_0 M_1 (t - t_0)^2 / 2(1 - \lambda)$ and e_j is a vector, which has all coordinates equal to 0 but j -th coordinate equal to 1. Component $i \neq j$ of this inequality is given by $0 \leq (t - t_0)[P_i(G + \zeta e_j) - P_i(G)] + \gamma_i(G, t)$. Dividing by $t - t_0 > 0$ and directing $t \rightarrow t_0$ on the right, considering

$\gamma_i(G, t)/(t - t_0) \rightarrow 0$ we get $0 \leq P(G + \zeta e_j) - P(G)$. Let us divide last expression by ζ and direct $\zeta \rightarrow 0$

$$0 \leq \lim_{\zeta \rightarrow 0^+} \frac{P(G + \zeta e_j) - P(G)}{\zeta} = \frac{\partial P_i}{\partial G_i} = J_{ij}$$

what is mean the fairing of statement (i). *End proof.*

Theorem 3. *Let Φ be Markovian mapping and $G^0, G^1 \in X$, $t \geq 0$, $\mu > 0$ than $\|z(t, G^1, \mu) - z(t, G^0, \mu)\| \leq \|G^1 - G^0\|$.*

Proof. Using (6) from the proofing of theorem 4 we have

$$\|z(t, G^1, \mu) - z(t, G^0, \mu)\| \leq \int_0^1 \|\Phi(t, \gamma(s))(G^1 - G^0)\| ds$$

Let function $\Phi(t, \gamma(s))$ is Markovian mapping for any

$$\|t \geq 0, s \in [0, 1], \|\Phi(t, \gamma(s))(G^1 - G^0)\| \leq \|G^1 - G^0\|.$$

Estimating the integral, considering this inequality, we get required. *End proof.*

This theorem shows us the following sufficient condition for the bound-
edness of the norm-solution $z(t, G, \mu)$.

Corollary fact from theorem 3. *Let $\exists G^* \in X: z(t, G^*, \mu) = G^*$.
Then $\|z(t, G, \mu) - G^*\| \leq \|G - G^*\|$ with $t \geq 0, G \in X$.*

This fact we can use for solutions analysis of the systems (7).

4. Conclusions

For the Cauchy problem for systems of differential equations of countable order, which describes the neural network with infinite number of cells, considered the question of the existence and uniqueness of its solution. Next step for investigation is constuting an algorithm for asymptotic solutions using approximate methods of solutions of differential equations with a small parameter at the highest derivative and analyzed the possibility of applying it solutions to predict the dynamics of complex systems in conditions of uncertainty.

Acknowledgments

The reported study was funded within the Agreement No 02.a03.21.0008 dated 24.11.2016 between the Ministry of Education and Science of the Russian Federation and RUDN University)

References

1. *Calvert B.D.* Neural networks with an infinite number of cells, *Journal of Differential Equations*. Vol. 186, Issue 1, pp. 31 – 51 (2002)
2. *Calvert B.D., Zemanian A.H.* Operating points in infinite nonlinear networks approximated by finite networks, *Trans. Amer. Math. Soc.* Vol. 352, No 2 , 753 – 780, (2000).
3. *Henry D.* Geometric theory of semilinear parabolic equations. *Lecture Notes in Mathematics*. Springer-Verlag, Berlin (1981)
4. *Hopfield J.J.* Neurons with graded response have collective computational properties like those of two state neurons, *Proc. Natl. Acad. Sci. USA*. Vol. 81, pp. 3088 – 3092 (1984)
5. *Haiying Huang, Qiaosheng Du, Xibing Kang.* Global exponential stability of neutral high-order stochastic Hopfield neural networks with Markovian jump parameters and mixed time delays. *ISA Transactions*. Vol. 52, Issue 6, pp. 759 – 767 (2013)
6. *Korobeinik Ju.* Differential equations of infinite order and infinite systems of differential equations. *Izv. Akad. Nauk SSSR Ser. Mat.* Vol. 34, pp. 881 – 922 (1970)
7. *Krasnoselsky M.A., Zabreyko P.P.* Geometrical methods of nonlinear analysis. Springer-Verlag, Berlin (1984)
8. *Lomov S. A.* The construction of asymptotic solutions of certain problems with parameters. *Izv. Akad. Nauk SSSR Ser. Mat.* Vol. 32, pp. 884 – 913 (1968)
9. *Persidsky K.P.* *Izv. AN KazSSR, Ser. Mat. Mach.*, Issue 2, pp. 3 – 34. (1948)
10. *Tihonov A. N.* Uber unendliche Systeme von Differentialgleichungen. *Rec. Math.* Vol. 41, Issue 4, pp. 551 – 555 (1934).
11. *Tihonov A. N.* Systems of differential equations containing small parameters in the derivatives. *Mat. Sbornik N. S.* Vol. 31, Issue 73, pp. 575 – 586 (1952)
12. *Vasil'eva A. B.* Asymptotic behaviour of solutions of certain problems for ordinary non-linear differential equations with a small parameter multiplying the highest derivatives. *Uspehi Mat. Nauk.* Vol. 18, Issue 111, no. 3 , 15 – 86 (1963)
13. *Zhautykov O. A.* On a countable system of differential equations with variable parameters. *Mat. Sb. (N.S.)*. Vol. 49, Issue 91, pp. 317 – 330 (1959)
14. *Zhautykov O. A.* Extension of the Hamilton-Jacobi theorems to an infinite canonical system of equations. *Mat. Sb. (N.S.)*. Vol. 53, Issue 95, pp. 313 – 328 (1961)
15. *Xiao Liang, Linshan Wang, Yangfan Wang, Ruili Wang.* Dynamical Behavior of Delayed Reaction-Diffusion Hopfield Neural Networks Driven by Infinite Dimensional Wiener Processes. *IEEE Transactions on Neural Networks*. Vol. 27, No 9, pp. 1816 – 1826 (2016)

UDC 004.4

Analysis of Numerical Solution Stability of Motion of Rolling Sphere on Rotating Plane

I. E. Kaspirovich

*Department of Theoretical Physics and Mechanics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

Abstract. In this paper we consider the problem of rolling sphere motion on the rotating plane. Using Euler angles as generalized coordinates numerical integration of these motion equations leads to the occurrence of singularities caused by degeneracy of this system of equations. Application of Rodrigo-Hamilton parameters allows not only to avoid these singularities but to simplify the procedure of numerical integration. In this paper the advantages of these parameters are demonstrated solving this problem. The condition of rolling without slipping constrains the system. Using numerical integration in this case leads to error accumulation caused by deviations from the constraints equations and, as a result, to the solution instability in relation to the constraints equations. Constraint stabilization can be applied to remove this instability using numerical integration. In this case given constraints are replaced with perturbed ones. Appropriate choice of perturbation parameters allows to avoid this error accumulation on each step of numerical integration and to achieve a stable solution.

Keywords: Constraint stabilization, Rodrigo - Hamilton parameters, Euler angles, Euler equations, Lagrange Multipliers.

1. Introduction

In this paper the problem of rolling sphere on a rotating plane is considered. The solution of this problem can be found in several ways. Detailed description of angles dependence on time or sphere's trajectory requires more complicated methods. For instance, when precession coincides with intrinsic rotation the system of motion equations in Euler angles is getting degenerated which leads to singularities in numerical integration.

2. Ordinary approach

Lagrange function of a unit mass sphere rolling without slipping on a rotating plane out of gravity field takes form

$$L = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + \frac{J}{2}(\omega_1^2 + \omega_2^2 + \omega_3^2), \quad (1)$$

where x, y - are center of mass coordinates (the density of a sphere is considered as a constant), J - is a moment of inertia about central axis, $\dot{x} = \frac{dx(t)}{dt}$ and $\vec{\omega} = (\omega_1, \omega_2, \omega_3)^T$ - is a vector of angular velocity.

The condition of rolling without slipping constraints the system with expressions:

$$\begin{aligned} f_1 &= \frac{d}{dt}x(t) - a\omega_2 + \Omega y(t) = 0, \\ f_2 &= \frac{d}{dt}y(t) + a\omega_1 - \Omega x(t) = 0, \end{aligned} \quad (2)$$

where a - is a radius of a sphere and Ω - is a rotation frequency of a plane.

One of the simplest but least informative method is solving Euler's equations for rigid body with inertia tensor $J_{ij}, i, j = 1, 2, 3$

$$J_{ij}\dot{\omega}_j + \varepsilon_{ijk}\omega_j J_{kl}\omega_l = 0. \quad (3)$$

Lets consider that the components of inertia tensor take form $J_{ij} = J\delta_{ij}, J = \frac{2}{5}a^2$ (δ_{ij} - is a Kronecker delta) then the equations with Lagrange multipliers u_1, u_2 for the equations (3) with constraints (2) are as follows:

$$\begin{cases} J\dot{\omega}_1 = au_2, \\ J\dot{\omega}_2 = -au_1, \\ J\dot{\omega}_3 = 0, \\ \ddot{x} = u_1, \\ \ddot{y} = u_2. \end{cases} \quad (4)$$

Equations (4) and constraints are exactly determined system. So Lagrange multipliers can be expressed through the generalized velocities and coordinates

$$\begin{aligned} u_1 &= -\frac{\Omega\dot{y}}{1 + \frac{a^2}{J}}, \\ u_2 &= \frac{\Omega\dot{x}}{1 + \frac{a^2}{J}}. \end{aligned}$$

Let $a = 1, \Omega = 1$, initial conditions $x(0) = 1, y(0) = 0, \dot{x}(0) = 0, \dot{y}(0) = 1, \omega_1(0) = 0, \omega_2(0) = 0, \omega_3(0) = 0$ then solving the system (4) the trajectory of center of mass appears to be an ellipse (see Fig. 1).

Euler equations for rigid body allow to obtain this trajectory, however, the angles' dependence on time remains unknown. It is necessary to add extra angular parameters, which can be expressed through angular velocities $\omega_i(\alpha, \beta, \gamma)$ in some differential form.

3. Rodrigo - Hamilton parameters

Euler angles θ, ψ, φ can be applied for a description of a rigid body rotation. Angular velocities $\omega_i, i = 1, 2, 3$ are expressed through them as

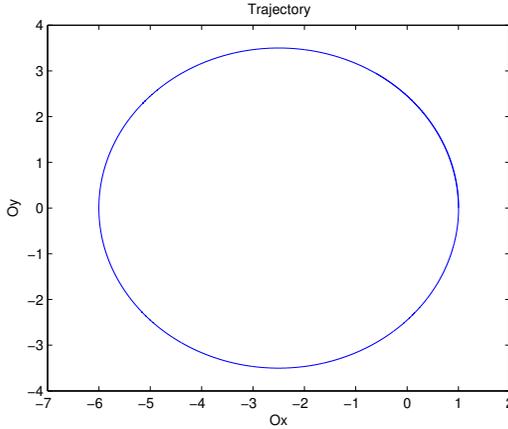


Figure 1. The trajectory of the center of the sphere.

quasi-coordinates in the following way:

$$\begin{aligned}\omega_1 &= \cos \psi \dot{\theta} + \sin \theta \sin \psi \dot{\varphi}, \\ \omega_2 &= \sin \psi \dot{\theta} - \sin \theta \cos \psi \dot{\varphi}, \\ \omega_3 &= \dot{\psi} + \cos \theta \dot{\varphi}.\end{aligned}$$

Lagrange function (1) in Euler angles takes form

$$L = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + \frac{J}{2}(\dot{\theta}^2 + \dot{\varphi}^2 + \dot{\psi}^2 + 2\dot{\varphi}\dot{\psi} \cos \theta).$$

However, there is a significant disadvantage in this approach. The system of motion equations in Euler angles is getting degenerated when the precession coincides with the intrinsic rotation:

$$\begin{aligned}\ddot{\theta} + \dot{\varphi}\dot{\psi} \sin \theta &= 0, \\ \ddot{\psi} + \ddot{\varphi} \cos \theta - \dot{\varphi}\dot{\theta} \sin \theta &= 0, \\ \ddot{\varphi} + \ddot{\psi} \cos \theta - \dot{\psi}\dot{\theta} \sin \theta &= 0.\end{aligned}$$

It is obvious, that if $\theta = \frac{\pi}{2}$ then the second and the third equation become identical. That is why the solvability of the second derivatives is impossible in this case. Rodrigo-Hamilton parameters [1] can be applied

to avoid this disadvantage. These parameters are components of some normalized quaternion $\Lambda = (\lambda_0, \vec{\lambda})$. Angular velocities ω_i are expressed through these components in a following way:

$$\begin{cases} \omega_1 = 2\lambda_0(t) \frac{d}{dt}\lambda_1(t) - 2\lambda_1(t) \frac{d}{dt}\lambda_0(t) + 2\lambda_3(t) \frac{d}{dt}\lambda_2(t) - 2\lambda_2(t) \frac{d}{dt}\lambda_3(t), \\ \omega_2 = 2\lambda_0(t) \frac{d}{dt}\lambda_2(t) - 2\lambda_2(t) \frac{d}{dt}\lambda_0(t) + 2\lambda_1(t) \frac{d}{dt}\lambda_3(t) - 2\lambda_3(t) \frac{d}{dt}\lambda_1(t), \\ \omega_3 = 2\lambda_0(t) \frac{d}{dt}\lambda_3(t) - 2\lambda_3(t) \frac{d}{dt}\lambda_0(t) + 2\lambda_2(t) \frac{d}{dt}\lambda_1(t) - 2\lambda_1(t) \frac{d}{dt}\lambda_2(t). \end{cases}$$

Normalizing condition $\lambda_0^2 + \vec{\lambda}^2 = 1$ allows to change Rodrigo-Hamilton parameters to some angular parameters α, β, γ :

$$\begin{aligned} \lambda_0 &= \cos \alpha \cos \beta \cos \gamma, \\ \lambda_1 &= \cos \alpha \cos \beta \sin \gamma, \\ \lambda_2 &= \cos \alpha \sin \beta, \\ \lambda_3 &= \sin \alpha. \end{aligned}$$

Angular velocities ω_i are expressed through α, β, γ in a following way:

$$\omega_i = l_i \dot{\gamma} + m_i \dot{\beta} + n_i \dot{\alpha}, \quad i = 1, 2, 3,$$

where l_i, m_i, n_i are as follows:

$$\begin{aligned} l_1 &= 2 \cos^2 \alpha \cos^2 \beta, & l_2 &= \cos^2 \alpha \sin 2\beta \sin \gamma - \sin 2\alpha \cos \beta \cos \gamma, \\ l_3 &= \cos^2 \alpha \sin 2\beta \cos \gamma + \sin 2\alpha \cos \beta \sin \gamma, & m_1 &= \sin 2\alpha \cos \beta, \\ m_2 &= \sin 2\alpha \sin \beta \sin \gamma + 2 \cos^2 \alpha \cos \gamma, & m_3 &= \sin 2\alpha \sin \beta \cos \gamma - 2 \cos^2 \alpha \sin \gamma, \\ n_1 &= -2 \sin \beta, & n_2 &= 2 \cos \beta \sin \gamma, & n_3 &= 2 \cos \beta \cos \gamma. \end{aligned} \quad (5)$$

Lagrange function (1) in α, β, γ takes form

$$L = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + 2J(\dot{\alpha}^2 + \cos^2 \alpha \dot{\beta}^2 + \cos^2 \alpha \cos^2 \beta \dot{\gamma}^2). \quad (6)$$

Lagrange function (6) looks like a Lagrange function for free particle on a three-dimensional sphere. In fact, a rotation of a circle around its centre generates a two-dimensional sphere, so the rotation of a sphere in four dimensions generates a three-dimensional one. One of the main advantage of angular parameters α, β, γ is absence of non-diagonal elements in Lagrange function (6). So, second derivatives can be solved for any values of these parameters.

Motion equations with Lagrange multipliers u_1, u_2 for Lagrange function (6) are as follows:

$$\begin{cases} \ddot{x} = u_1, \\ \ddot{y} = u_2, \\ \ddot{\alpha} = -\frac{1}{2} \sin 2\alpha (\dot{\beta}^2 + \cos^2 \beta \dot{\gamma}^2) + \frac{a}{4J} (u_2 n_1 - u_1 n_2), \\ \ddot{\beta} = \frac{1}{2} \sin 2\beta \dot{\gamma}^2 + 2 \operatorname{tg} \alpha \dot{\alpha} \dot{\beta} + \frac{a}{4J \cos^2 \alpha} (u_2 m_1 - u_1 m_2), \\ \ddot{\gamma} = 2 \operatorname{tg} \alpha \dot{\alpha} \dot{\gamma} + 2 \operatorname{tg} \beta \dot{\beta} \dot{\gamma} + \frac{a}{4J \cos^2 \alpha \cos^2 \beta} (u_2 l_1 - u_1 l_2). \end{cases} \quad (7)$$

The system of equations (7), (5) and the constraint equation (2) allow to express multipliers as functions of coordinates and velocities. Euler method is applied for numerical integration of in this paper. However, this system also contains singularities when $\alpha = \frac{\pi}{2}$, $\beta = \frac{\pi}{2}$, but to obtain the angular dependence on time singularity points can be skipped in a cycle of numerical integration in a following way:

$$\text{if } \beta_i \in \left(\frac{\pi n}{2} - \delta, \frac{\pi n}{2} + \delta \right) \text{ then } \beta_{i+1} = \beta_i + \varepsilon, \quad n \in \mathbb{Z},$$

where δ and ε are infinitesimals. This approach allows to avoid infinities in numerical integration, but it causes massive deviations from the solution. The method of constraint stabilization [2] allows to minimize these deviations. In this method constraints f_i are replaced with perturbed ones $\dot{f}_i = k_{ij} f_j$, where k_{ij} are constants. For example, after the first singularity $\beta = \frac{\pi}{2}$, if the perturbation parameters take form $k_{11} = 0$, $k_{12} = -10$, $k_{21} = 0$, $k_{22} = 0$, then in numerical integration deviations are smaller than they would be if all these parameters equal to zero (see Fig. 2). However, after several skips the method of constraint stabilization is getting less effective (see Fig. 3).

This plot β looks like the stairs with the stages on the site of skips. Bigger amount of skips leads to the higher stage, so the trajectory after 5 skips is presented in Fig. 4.

4. Conclusions

It is quite enough to get a trajectory of the sphere just solving Euler equations. But to get more information about angular dependence we need to apply some extra methods. The method of skipping singularities allows to avoid infinities in numerical integration and constraint stabilization allows not only to get a stable numerical solution but to minimize the deviations caused by these skips.

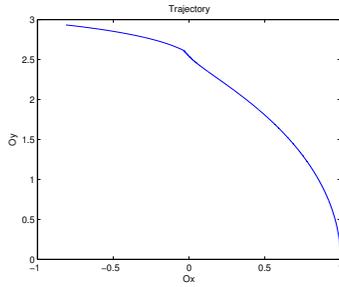


Figure 2. The deviation of the trajectory after the first skip

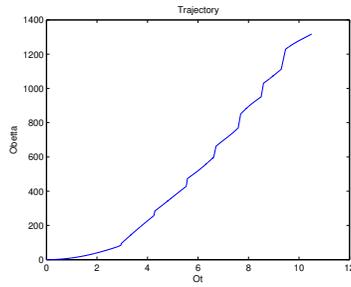


Figure 3. Dependence β on time

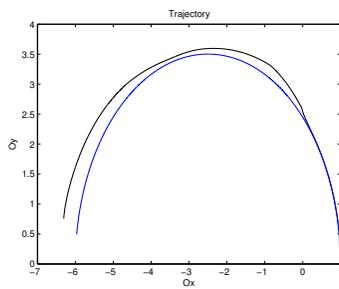


Figure 4. The trajectory of the center of the sphere after 5 skips.

Acknowledgments

This work was supported by grant 16-08-00558 from the Russian Foundation for Basic Research and by government task 3.1939.2014/K; project part.

References

1. *Amelkin N. I.* Rigid Body Kinematics and Dynamics — Moscow, 2000, —
2. *Mukharlyamov R. G.* Simulation of Control Processes, Stability and Stabilization of Systems with Program Constraints. — in Journal of Computer and Systems Sciences International, vol. 54, pp. 13-26, 2015., 1994.

UDC 004.4

Low-priority queue and server's steady-state existence in a tandem under prolongable cyclic service

V. M. Kocheganov*[†], A. V. Zorine*

* *Department of Applied Probability Theory
N. I. Lobachevsky State University of Nizhni Novgorod
Gagarina av. 23, 603950 Nizhny Novgorod, Russia*

[†] *Department of Applied Probability Theory
N. I. Lobachevsky State University of Nizhni Novgorod
Gagarina av. 23, 603950 Nizhny Novgorod, Russia*

Abstract. A mathematical model of a tandem of queuing systems is considered. Each system has a high-priority input flow and a low-priority input flow which are conflicting. In the first system, the customers are serviced in the class of cyclic algorithms. The serviced high-priority customers are transferred from the first system to the second one with random delays and become the high-priority input flow of the second system. In the second system, customers are serviced in the class of cyclic algorithms with prolongations. Low-priority customers are serviced when their number exceeds a threshold. A mathematical model is constructed in form of a multidimensional denumerable discrete-time Markov chain. Conditions of low-priority queue stationary distribution existence were found.

Keywords: tandem of controlling queuing systems, cyclic algorithm with prolongations, conflicting flows, multidimensional denumerable discrete-time Markov chain.

1. Introduction

An enormous amount of work has been done on the problem of conflicting traffic flows control at crossroad by the moment. In the queuing theory literature one can find following algorithms investigated: fixed duration cyclic algorithm, cyclic algorithm with a loop, cyclic algorithm with changing regimes, etc [1–6]. However, in a real-life situations cars pass several consecutive crossroads on their way rather than only one. In other words, an output flow of cars from the first intersection forms an input flow of cars of the next intersection. Hence, the second input flow no longer has an *a priori* known simple probabilistic structure (for example, a non-ordinary Poisson flow), and knowledge about the service algorithm should be taken into account to deduce formation conditions of the first output flow.

One can find several works about tandems of intersections. In [7] a computer-aided simulation of adjacent intersections was carried out. In [8] a mathematical model of two intersections in tandem governed by cyclic algorithms was investigated and stability conditions were found. In this paper we assume that the first intersection is governed by a cyclic algorithm while the second intersection is governed by a cyclic algorithm with

prolongations. The low-priority queue in the second intersection and necessary conditions of its stationary state existence take central place of this paper. This work continues studying in paper [10].

2. The problem settings

Consider a queuing system with a scheme shown in Fig. 1. There are four input flows of customers $\Pi_1, \Pi_2, \Pi_3,$ and Π_4 entering the single server queueing system. Customers in the input flow $\Pi_j, j \in \{1, 2, 3, 4\}$ join a queue O_j with an unlimited capacity. For $j \in \{1, 2, 3\}$ the discipline of the queue O_j is FIFO (First In First Out). Discipline of the queue O_4 will be described later. The input flows Π_1 and Π_3 are generated by an external environment, which has only one state. Each of these flows is a nonordinary Poisson flow. Denote by λ_1 and λ_3 the intensities of bulk arrivals for the flows Π_1 and Π_3 respectively. The probability generating function of number of customers in a bulk in the flow Π_j is

$$f_j(z) = \sum_{\nu=1}^{\infty} p_{\nu}^{(j)} z^{\nu}, \quad j \in \{1, 3\}. \quad (1)$$

We assume that $f_j(z)$ converges for any $z \in \mathbb{C}$ such that $|z| < (1 + \varepsilon)$, $\varepsilon > 0$. Here $p_{\nu}^{(j)}$ is the probability of a bulk size in flow Π_j being exactly $\nu = 0, 1, \dots$. Having been serviced the customers from O_1 come back to the system as the Π_4 customers. The Π_4 customers in turn after service enter the system as the Π_2 ones. The flows Π_2 and Π_3 are conflicting in the sense that their customers can't be serviced simultaneously. This implies that the problem can't be reduced to a problem with fewer input flows by merging the flows together.

In order to describe the server behavior we fix positive integers d, n_0, n_1, \dots, n_d and we introduce a finite set $\Gamma = \{\Gamma^{(k,r)} : k = 0, 1, \dots, d; r = 1, 2, \dots, n_k\}$ of states server can reside in. At the state $\Gamma^{(k,r)}$ server stays during constant time $T^{(k,r)}$. Define disjoint subsets $\Gamma^I, \Gamma^{II}, \Gamma^{III},$ and Γ^{IV} of Γ as follows. In the state $\gamma \in \Gamma^I$ only customers from the queues O_1, O_2 and O_4 are serviced. In the state $\gamma \in \Gamma^{II}$ only customers from the queues O_2 and O_4 are serviced. In the state $\gamma \in \Gamma^{III}$ only customers from queues $O_1, O_3,$ and O_4 are serviced. In the state $\gamma \in \Gamma^{IV}$ only customers from queues O_3 and O_4 are serviced. We assume that $\Gamma = \Gamma^I \cup \Gamma^{II} \cup \Gamma^{III} \cup \Gamma^{IV}$. Set also ${}^1\Gamma = \Gamma^I \cup \Gamma^{III}, {}^2\Gamma = \Gamma^I \cup \Gamma^{II}, {}^3\Gamma = \Gamma^{III} \cup \Gamma^{IV}$.

The server changes its state according to the following rules. We call a set $C_k = \{\Gamma^{(k,r)} : r = 1, 2, \dots, n_k\}$ the k -th cycle, $k = 1, 2, \dots, d$. For $k = 0$ the state $\Gamma^{(0,r)}$ with $r = 1, 2, \dots, n_0$ is called a prolongation state. Put $r \oplus_k 1 = r + 1$ for $r < n_k$, and $r \oplus_k 1 = 1$ for $r = n_k$ ($k = 0, 1, \dots, d$). In the cycle C_k we select a subset C_k^O of input states, a subset

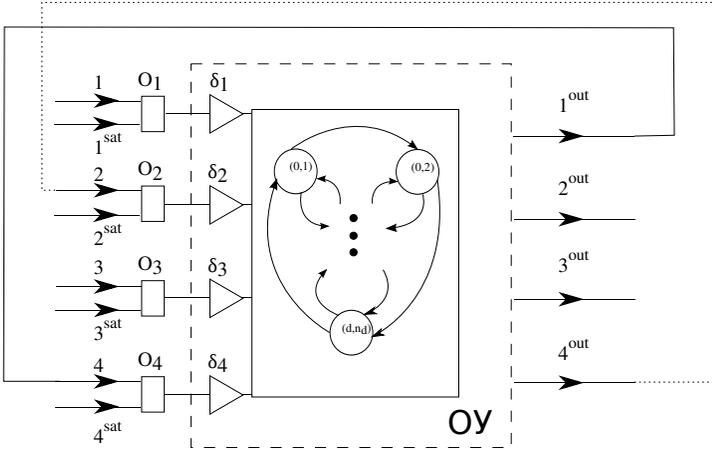


Figure 1. Scheme of the queuing system as a cybernetic control system

C_k^I of output states, and a subset $C_k^N = C_k \setminus (C_k^O \cup C_k^I)$ of neutral states. After the state $\Gamma^{(k,r)} \in C_k \setminus C_k^O$ the server switches to the state $\Gamma^{(k,r \oplus_k 1)}$ within the same cycle C_k . After the state $\Gamma^{(k,r)}$ in C_k^O the server switches to the state $\Gamma^{(0,r \oplus_k 1)}$ if number of customers in the queue O_3 at switching instant is greater than a predetermined threshold L . Otherwise, if the number of customers in the queue O_3 is less than or equals L then the new state is the prolongation one $\Gamma^{(0,r_1)}$ where $r_1 = h_1(\Gamma^{(k,r)})$ and $h_1(\cdot)$ is a given mapping of $\bigcup_{k=1}^d C_k^O$ into $\{1, 2, \dots, n_0\}$. After the state $\Gamma^{(0,r)}$ if the number of customers in O_3 is not above L the state of the same type $\Gamma^{(0,r_2)}$ is chosen where $r_2 = h_2(r)$ and $h_2(\cdot)$ is a given mapping of the set $\{1, 2, \dots, n_0\}$ into itself; in the other case the new state is $\Gamma^{(k,r_3)} \in C_k^I$ where $\Gamma^{(k,r_3)} = h_3(r)$ and $h_3(\cdot)$ is a given mapping of $\{1, 2, \dots, n_0\}$ to $\bigcup_{k=1}^d C_k^I$. We assume that each prolongation state $\Gamma^{(0,r)}$ belongs to the set ${}^2\Gamma$ and that relations $C_k^O \subset {}^2\Gamma$ and $C_k^I \subset {}^3\Gamma$ hold. We also assume that all the cycles have exactly one input and output state. Finally, we assume that all the prolongation states make a cycle, that is $h_2(r) = r \oplus_0 1$. Putting all together, we introduce a function which formalizes the server state changes (see equation (2)).

In general, service durations of different customers can be dependent and may have different laws of probability distributions. So, saturation flows will be used to define the service process. A saturation flow

Π_j^{sat} , $j \in \{1, 2, 3, 4\}$, is defined as a virtual output flow under the maximum usage of the server and unlimited number of customer in the queue O_j . The saturation flow Π_j^{sat} , $j \in \{1, 2, 3\}$ contains a non-random number $\ell(k, r, j) \geq 0$ of customers in the server state $\Gamma^{(k,r)}$. In particular, $\ell(k, r, j) \geq 1$ for $\Gamma^{(k,r)} \in j\Gamma$ and $\ell(k, r, j) = 0$ for $\Gamma^{(k,r)} \notin j\Gamma$. Let \mathbb{Z}_+ be the set of non-negative integer numbers. If the queue O_4 contains $x \in \mathbb{Z}_+$ customers the saturation flow Π_4^{sat} also contains the x customers. Finally, in the state $\Gamma^{(k,r)}$ every customer from queue O_4 with probability $p_{k,r}$ and independently of others ends servicing and joins Π_2 to go to O_2 . With the complementary probability $1 - p_{k,r}$ the customer stays in O_4 until the next time slot. In the next time slot it repeats its attempt to join Π_2 with a proper probability.

$$h(\Gamma^{(k,r)}, y) = \begin{cases} \Gamma^{(k,r \oplus_k 1)} & \text{if } \Gamma^{(k,r)} \in C_k \setminus C_k^O \text{ or} \\ & (\Gamma^{(k,r)} \in C_k^O) \wedge (y > L); \\ \Gamma^{(0, h_1(\Gamma^{(k,r)}))} & \text{if } \Gamma^{(k,r)} \in C_k^O \text{ and } y \leq L; \\ \Gamma^{(0, r \oplus_{01})} & \text{if } k = 0 \text{ and } y \leq L; \\ h_3(r) & \text{if } k = 0 \text{ and } y > L. \end{cases} \quad (2)$$

A real-life example of just described queuing system is a tandem of two consecutive crossroads (Fig. 2). The input flows are flows of vehicles.

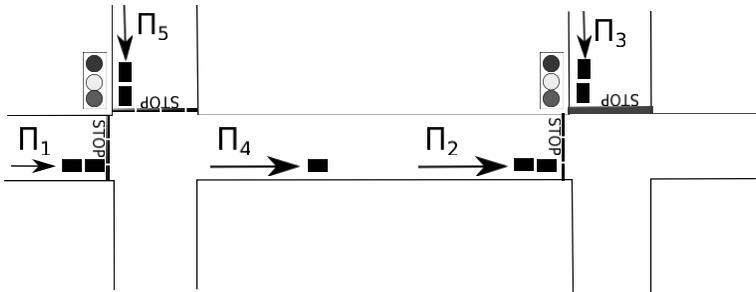


Figure 2. A tandem of crossroads, the physical interpretation of the queuing system under study

The flows Π_1 and Π_5 at the first crossroad are conflicting; Π_2 and Π_3 at the second crossroad are also conflicting. Every vehicle from the flow Π_1 after passing first road intersection joint the flow Π_4 and enters the queue

O_4 . After some random time interval the vehicle arrives to the next road intersection. Such a pair of crossroads is an instance of a more general queuing model described above.

3. Mathematical model

The queuing system under investigation can be regarded as a cybernetic control system that helps to rigorously construct a formal stochastic model [8]. The scheme of the control system is shown in Fig. 1. There are following blocks present in the scheme: 1) the external environment with one state; 2) input poles of the first type — the input flows Π_1, Π_2, Π_3 , and Π_4 ; 3) input poles of the second type — the saturation flows $\Pi_1^{\text{sat}}, \Pi_2^{\text{sat}}, \Pi_3^{\text{sat}}$, and Π_4^{sat} ; 4) an external memory — the queues O_1, O_2, O_3 , and O_4 ; 5) an information processing device for the external memory — the queue discipline units $\delta_1, \delta_2, \delta_3$, and δ_4 ; 6) an internal memory — the server (OY); 7) an information processing device for internal memory — the graph of server state transitions; 8) output poles — the output flows $\Pi_1^{\text{out}}, \Pi_2^{\text{out}}, \Pi_3^{\text{out}}$, and Π_4^{out} . The coordinate of a block is its number on the scheme.

Let us introduce the following variables and elements along with their value ranges. To fix a discrete time scale consider the epochs $\tau_0 = 0, \tau_1, \tau_2, \dots$ when the server changes its state. Let $\Gamma_i \in \Gamma$ be the server state during the interval $(\tau_{i-1}; \tau_i]$, $\varkappa_{j,i} \in \mathbb{Z}_+$ be the number of customers in the queue O_j at the instant τ_i , $\eta_{j,i} \in \mathbb{Z}_+$ be the number of customers arrived into the queue O_j from the flow Π_j during the interval $(\tau_i; \tau_{i+1}]$, $\xi_{j,i} \in \mathbb{Z}_+$ be the number of customers in the saturation flow Π_j^{sat} during the interval $(\tau_i; \tau_{i+1}]$, $\bar{\xi}_{j,i} \in \mathbb{Z}_+$ be the actual number of serviced customers from the queue O_j during the interval $(\tau_i; \tau_{i+1}]$, $j \in \{1, 2, 3, 4\}$.

The server changes its state according to the following rule:

$$\Gamma_{i+1} = h(\Gamma_i, \varkappa_{3,i})$$

where the mapping $h(\cdot, \cdot)$ is defined by Formula (2).

Let $\varphi_1(\cdot, \cdot)$ and $\varphi_3(\cdot, \cdot)$ be defined by series expansions

$$\sum_{\nu=0}^{\infty} z^{\nu} \varphi_j(\nu, t) = \exp\{\lambda_j t (f_j(z) - 1)\}$$

with functions $f_j(z)$ defined by (1), $j \in \{1, 3\}$. The function $\varphi_j(\nu, t)$ equals the probability of $\nu = 0, 1, \dots$ arrivals in the flow Π_j during time $t \geq 0$. If $\nu < 0$ the value of $\varphi_j(\nu, t)$ is set to zero.

Mathematical model in more details can be found in work [10]. From now on we focus on low-priority customers in the queue O_3 .

4. The low-priority queue

Here we will consider the stochastic sequence

$$\{(\Gamma_i(\omega), \varkappa_{3,i}(\omega)); i = 0, 1, \dots\}, \tag{3}$$

which includes the number of low-priority customers $\varkappa_{3,i}(\omega)$ in the queue O_3 . In this section we will report several results concerning this stochastic sequence.

Let $\Gamma^{(k,r)} \in \Gamma$ and $x_3 \in Z_+$. Denote by $H_{-1}(\Gamma^{(k,r)}, x_3)$ the set of all server states γ such that $h(\gamma, x_3) = \Gamma^{(k,r)}$ and put $r \ominus_k 1 = r - 1$ for $n_k \geq r > 0$, and $r \ominus_k 1 = n_k$ for $r = 0$ ($k = 0, 1, \dots, d$). Then formula (2) makes it possible to define the mapping $H_{-1}(\Gamma^{(k,r)}, x_3)$ explicitly:

$$H_{-1}(\Gamma^{(k,r)}, x_3) = \begin{cases} \{\Gamma^{(k_1, r_1)}, \Gamma^{(0, r \ominus_0 1)}\} & \text{if } (k = 0) \wedge (x_3 \leq L), \\ \{\Gamma^{(k, r \ominus_k 1)}, \Gamma^{(0, r_2)}\} & \text{if } (\Gamma^{(k,r)} \in C_k^I) \wedge (x_3 > L), \\ \{\Gamma^{(k, r \ominus_k 1)}\} & \text{if } (\Gamma^{(k,r)} \in C_k^O) \vee (\Gamma^{(k,r)} \in C_k^N); \\ \emptyset & \text{if } (k = 0) \wedge (x_3 > L) \\ & \text{or } (\Gamma^{(k,r)} \in C_k^I) \wedge (x_3 \leq L) \end{cases}$$

where $h_1(\Gamma^{(k_1, r_1)}) = r$ and $h_3(r_2) = \Gamma^{(k,r)}$.

Let's define for $\gamma \in \Gamma$ and $x_3 \in Z_+$ values

$$Q_{3,i}(\gamma, x) = \mathbf{P}(\{\omega : \Gamma_i(\omega) = \gamma, \varkappa_{3,i}(\omega) = x_3\}).$$

Theorem 1 concerns generating functions and corrects ones in paper [10]. Suppose k and r are such that $\Gamma^{(k,r)} \in \Gamma$. Let's define partial probability generating functions

$$\mathfrak{M}^{(i)}(k, r, v) = \sum_{w=0}^{\infty} Q_{3,i}(\Gamma^{(k,r)}, w)v^w,$$

$$q_{k,r}(v) = v^{-\ell(k,r,3)} \sum_{w=0}^{\infty} \varphi_3(w, T^{(k,r)})v^w.$$

and auxiliary functions

$$\tilde{\alpha}_i(\tilde{k}, \tilde{r}, v) = \sum_{x_3=0}^{\ell(\tilde{k}, \tilde{r}, 3)} \sum_{\gamma \in H_{-1}(\tilde{\gamma}, x_3)} Q_{3,i}(\gamma, x_3) \sum_{a=0}^{\ell(\tilde{k}, \tilde{r}, 3) - x_3} \varphi_3(a, T^{(\tilde{k}, \tilde{r})}) -$$

$$- \sum_{x_3=0}^{\ell(\tilde{k}, \tilde{r}, 3)} \sum_{\gamma \in H_{-1}(\tilde{\gamma}, x_3)} Q_{3,i}(\gamma, x_3) v^{x_3 - \ell(\tilde{k}, \tilde{r}, 3)} \sum_{w=0}^{\ell(\tilde{k}, \tilde{r}, 3) - x_3} \varphi_3(w, T^{(\tilde{k}, \tilde{r})}) v^w$$

$$\alpha_i(0, \tilde{r}, v) = \tilde{\alpha}_i(0, \tilde{r}, v) + q_{0, \tilde{r}}(v) \times \sum_{x_3=0}^L \left[Q_{3,i}(\Gamma^{(k_1, r_1)}, x_3) + Q_{3,i}(\Gamma^{(0, r \ominus 01)}, x_3) \right] v^{x_3}, \quad \Gamma^{(0, \tilde{r})} \in \Gamma.$$

$$\alpha_i(\tilde{k}, \tilde{r}, v) = \tilde{\alpha}_i(\tilde{k}, \tilde{r}, v) - q_{\tilde{k}, \tilde{r}}(v) \sum_{x_3=0}^L \left[Q_{3,i}(\Gamma^{(\tilde{k}, \tilde{r} \ominus \tilde{k}1)}, x_3) + Q_{3,i}(\Gamma^{(0, r_2)}, x_3) \right] v^{x_3} + q_{\tilde{k}, \tilde{r}}(v) \sum_{x_3 \geq 0} Q_{3,i}(\Gamma^{(0, \tilde{r})}, x_3) v^{x_3}, \quad \Gamma^{(\tilde{k}, \tilde{r})} \in C_{\tilde{k}}^I$$

$$\alpha_i(\tilde{k}, \tilde{r}, v) = \tilde{\alpha}_i(\tilde{k}, \tilde{r}, v), \quad \Gamma^{(\tilde{k}, \tilde{r})} \in C_{\tilde{k}}^O \cup C_{\tilde{k}}^N.$$

Theorem 1 *Following recurrent w.r.t. $i \geq 0$ relations take place for the partial probability generating functions:*

1. $\Gamma^{(0, \tilde{r})} \in \Gamma, \tilde{r} = \overline{1, n_0}$

$$\mathfrak{M}^{(i+1)}(0, \tilde{r}, v) = \alpha_i(0, \tilde{r}, v);$$

2. $\Gamma^{(\tilde{k}, \tilde{r})} \in \Gamma, \tilde{k} = \overline{1, d}, \tilde{r} = \overline{1, n_{\tilde{k}}}$

$$\mathfrak{M}^{(i+1)}(\tilde{k}, \tilde{r}, v) = q_{\tilde{k}, \tilde{r}}(v) \times \mathfrak{M}^{(i)}(\tilde{k}, \tilde{r} \ominus_{\tilde{k}} 1, v) + \alpha_i(\tilde{k}, \tilde{r}, v);$$

The last result (theorem 2) is new and concerns low-priority queue and server’s steady-state existence.

Theorem 2 *For Markov chain (3) to have stationary distribution $Q_3(\gamma, x), (\gamma, x) \in \Gamma \times \mathbb{Z}_+$ it is necessary that following inequality takes place*

$$\min_{k=1, d} \frac{\lambda_3 f'_3(1) \sum_{r=1}^{n_k} T^{(k, r)}}{\sum_{r=1}^{n_k} \ell(k, r, 3)} < 1.$$

5. Acknowledgments

This work was fulfilled as a part of State Budget Research and Development program No. 01201456585 “Mathematical modeling and analysis of stochastic evolutionary systems and decision processes” of N.I. Lobachevsky State University of Nizhni Novgorod and was supported by State Program “Promoting the competitiveness among world’s leading research and educational centers”

References

1. Neimark Yu. I., Fedotkin M. A., Preobrazhenskaja A. M. Operation of an automate with feedback controlling traffic at an intersection // *Izvestija of USSR Academy of Sciences, Technical Cybernetic*. 1968. No. 5. P. 129–141.
2. Fedotkin M. A. On a class of stable algorithms for control of conflicting flows or arriving airplanes // *Problems of control and information theory*. 1977. V. 6, No. 1. P. 13–22.
3. Fedotkin M. A. Construction of a model and investigation of nonlinear algorithms for control of intense conflict flows in a system with variable structure of servicing demands. I // *Lithuanian mathematical journal*. 1977. V. 17, No. 1. P. 129–137.
4. Litvak N. V., Fedotkin M. A. A probabilistic model for the adaptive control of conflict flows // *Automation and Remote Control*. 2000. V. 61, No. 5. P. 777–784.
5. Proidakova E.V., Fedotkin M.A. Control of output flows in the system with cyclic servicing and readjustments // *Automation and remote control*. 2008. V. 69, No. 6. P. 993–1002.
6. Afanasyeva L. G., Bulinskaya E. V. Mathematical models of transport systems based on queueing theory // *Trudy of Moscow Institute of Physocs and Technology*. 2010. No. 4. P.6–21.
7. Yamada K., Lam T. N. Simulation analysis of two adjacent traffic signals // *Proceedings of the 17th winter simulation conference*. ACM, New York. 1985. P. 454–464.
8. Zorin A.V. Stability of a tandem of queueing systems with Bernoulli noninstantaneous transfer of customers // *Theory of Probability and Mathematical Statistics*. 2012. V. 84. P. 173–188.
9. Kocheganov V. M., Zorine A.V. Probabilistic model of tandem of queueing systems under cyclic control with prolongations // *Proceedings of Internation conference “Probability theory, stochastic processes, mathematical statistics and applications” (Minsk, Feb. 23–26 2015)*. 2015. P. 94–99.
10. Kocheganov V.M., Zorine A.V. Low-Priority Queue Fluctuations in Tandem of Queueing Systems Under Cyclic Control with Prolongations // *Distributed Computer and Communication Networks. Ser. Communications in Computer and Information Science*. 2016. V. 601. P. 268–279.

Analysis of the effectiveness of forward error correction in selective mode of transport protocol

V. Kokshenev[†], P. Mikheev*, S. Suschenko*, R. Tkachev*

* *Tomsk State University
Lenina 36, Tomsk, 634050, Russia*

[†] *F5 Networks Inc.. (USA)*

Abstract. The proposed model is a virtual connection managed by the transport protocol with a forward error correction mechanism for selective repeat mode in the form of Markov chain with discrete time. The analysis of the impact of protocol parameters window size and the duration of the timeout of waiting confirmation, the likelihood of distortion of the segments in the individual links of the transmission path data, the duration of the round-trip delay, the parameters of mechanism to restore the distorted segments (without retransmissions) on throughput of a transport connection. In the area of protocol parameters, the characteristics of the transmission channel and parameters of the forward error correction mechanism found in the area of superiority of the management procedures of the transport protocol with forward error correction over the classic procedure with decision feedback on the criterion of the throughput of a transport connection. The expediency of applying of the method of forward error correction for transport links with large round-trip delay.

Keywords: transport protocol, data path, forward error correction, Markov chain, throughput of a transport connection, window size, duration of the timeout, round-trip delay, loss rate.

1. Introduction

The most important indicator of the quality of interaction and networking applications used software and hardware of computer networks is the throughput transport links. This operating characteristic is largely determined by the transport protocol, its parameters — the width of the window and the duration of the timeout [1,2], as well as additional mechanisms to increase performance by reducing the number of retransmissions of distorted data [3–5]. Simulation of the subscriber connection and the analysis of its potential ability to perform is in the [2–10], and other works. But the results were obtained only for single-link data path [6–8], or with significant restrictions on the protocol parameters [9,10]. Known technology of forward error correction (FEC) are used as the transport protocol typically as services for the lower level network architecture [3–5]. A comprehensive analysis of the advantages and effectiveness of the methods of advanced error correction performed only on a qualitative level, as well as numerically for a number of special cases and not allowed to identify areas of possible application methods in the area of protocol parameters and characteristics of the transport connection. Modern transport protocols contain a wide variety of congestion control mechanisms [11]. There is a wide range of studies [11,12] in the field of control parameters of the

transport protocol with the aim of preventing and bypass congestion, focused on building models of diagnostic over various indicators [11] and adaptation of protocol parameters and mechanisms of error correction to changing network load and connectivity, the level of losses, activity interactive subscribers, etc. conditions of data transmission. However, the implementation of control mechanisms to bypass congestion based on available bandwidth of the transport connection at the current and predictable to changing the values of protocol parameters. Thus, the potential of the transport protocol using the methods of forward error correction has not been studied yet. There is no analytical dependence of the overall effect of protocol parameters, the characteristics of the transmission channel data and parameters of the correction method, the resulting operating characteristics of the transport connection. Not studied the effect of correlations between the duration of round-trip delay and protocol parameters on the throughput of the transmission path of the managed data transport protocol. In addition, the process of data transmission in computer networks are essentially discrete in nature [13], due to the pipeline transfer mechanism in the network of limited size segments and the use of algorithms with decision feedback at different levels of network architecture, however, most of the results [2–7, 11] based on models with continuous time, which leads to the narrowing of the field of their applicability.

The paper presents a mathematical model of the data transfer process from the FEC in the phase of information transfer in the form of a Markov chain with discrete time, analytically found stationary distribution of probabilities of states for the mode of selective reject, [2] obtained analytical expressions for the bandwidth on the basis of which the analysis of potential capabilities of the transport connection is performed.

2. Transport connection model

Let's consider the process of transferring data between subscribers of transport protocol based on the algorithm with decision feedback and operating in a selective reject mode [2]. An example of such reliable family of protocols is dominant in modern computer networks, the TCP protocol [1]. We believe that interactive subscribers have unlimited data stream for transmission, and the exchange is performed by data units of the transport protocol (segments) of equal length. We believe that the area of hops along the transmission path data have the same speed in both directions, and the length of loop segment in a separate link is t . In general, the path length from the source to the destination of the transfer information flow, and the length of the return path that are transmitted to received acknowledgment segments can be different. We believe that the data length of the transmission path, expressed in hop number of sites in the forward direction is equal $D_n \geq 1$. Reverse path delivers the confirmation to the sender about the correct reception of the blocks of the sequence of segments has

length $D_o \geq 1$. Set the probability of segment distortions in communication channels for forward $R_n(d)$, $d = 1, \overline{D}_n$ and reverse $R_o(d)$, $d = 1, \overline{D}_o$ directions of transmission of each part of hops. Then the reliability of segments transmission along the path from the source to the destination and back will be $F_n = \prod_{d=1}^{\overline{D}_n} (1 - R_n(d))$ and $F_o = \prod_{d=1}^{\overline{D}_o} (1 - R_o(d))$ respectively. We believe that the loss of segments due to the absence in the buffer memory in tract's nodes does not occur. We believe that data transmission by the sender is implemented by blocks containing B segments of which $1 \leq A \leq B$ — are informative, and $B - A$ — additional (redundant) of the same length. Loss (distortion) to $B - A$ arbitrary segments in the block allows you to restore all of the block segments for RAID-arrays of the fourth level [14]. Flow control mechanism is implemented by sliding window [1, 2] protocol parameters with a window width $w \geq 1$, expressed in a number of blocks. We believe that the verification of the correctness of the received target segments blocks are transferred in each segment of the oncoming flow. In case of impossibility of direct recovery of the transmitted segment blocks (the distortion more $B - A$ segments in the block) entire block is retransmitted. Then, the process of information transfer in virtual connection managed by the transport protocol, can be described by Markov process with discrete time (with the cycle duration t) due to the fact that the time between the receipt of confirmation has the geometric distribution with parameter F_o . This model is a generalization of the formalizations of the data transfer process, proposed in [3–6], in case of a transport connection of arbitrary length and the mechanism of forward error correction. The area of possible states of the Markov's chain determined by the duration of the timeout of waiting confirmation S , expressed in number of cycles of duration t . The size of the time-out is associated with the length of the tract, the width of the window and a block size of the inequations $S \geq wB + 1$, $S \geq \overline{D}_n + \overline{D}_o + B - 1$. It is obvious that the sum of the lengths of the forward and backward paths can be interpreted as a circular delay of a single segment $D = \overline{D}_n + \overline{D}_o$, expressed in the durations t (excluding losses of protocol blocks in the transmission along the path). The cyclic delay for block of segments will be $D + B - 1$. The states of the Markov chain $i = 0, w\overline{B}$ corresponds to the size of the queue transmitted, but not confirmed segments in the stream, and the states $i = \overline{wB + 1}, \overline{S - 1}$ — the time during which the sender is not active and is waiting to receive confirmation about the correctness of receiving the transmitted sequence from w blocks of segments. From the zero state in the $(D + B - 2)$ -th sender is moving with every cycle t with a probability of determine event. In the states $i \geq D + B - 2$ after the expiration of the next discrete cycle t the sender begins to receive confirmation and, depending on the results of the delivery units of the segments according to the technology of forward error correction, the sender transmits new blocks of segments (with positive confirmation) or re-distorted (not allowing to direct recovery). The completion of the cycle of the host state $D + B - 2$

corresponds to the time of bringing the first set of segments to the destination and receiving an acknowledgment. Further growth of the state occurs with a probability of distortion of the confirmation $1 - F_o$ in the reverse path. In states $i \geq D + B - 2$ in the selective reject mode the confirmation gives conversion to $(D - 1)$ -th state when $w \geq K + 2$, $K = \lfloor \frac{D-2}{B} \rfloor$, where $\lfloor \dots \rfloor$ — means “integer part” of the fraction, or in the state kB , $k = \overline{1, w - 1}$ where $w \leq K + 1$. Due to the fact that in states $i \geq wB$ the source stops sending blocks of segments, obtaining confirmation when $w \geq K + 2$ in the states $i = \overline{(w + 1)B - 1, (w + K + 1)B - 2}$, $k = \overline{1, K}$ leads to a transition in state $D - kB - 1$, $k = \overline{1, K}$. When $w \leq K + 1$ out of state $i = \overline{D + (w - K)B - 2, D + (w - k + 1)B - 3}$, $k = \overline{1, K}$ there is a transition to state kB , $k = \overline{1, w - 1}$. In states $i = \overline{(w + K + 1)B - 1, S - 2}$ with arbitrary width of the window there is a transition to zero state, since the size of the queue transmitted but unconfirmed information segments come clear. In the state $S - 1$ the timeout period of waiting for acknowledgment from the recipient runs out, of the correctness of the received blocks and segments and there is unconditional conversion to zero state.

3. The state probabilities for selective rejection mode with forward error correction mechanism

The transition probabilities π_{ij} from the initial state i in the resulting j Markov's chain that describes the process of transmission of information flow with the technology of direct correction error in selective reject mode have the form:

$$\pi_{ij} = \begin{cases} 1, & j = i + 1, i = \overline{0, D + B - 3}; \\ 1 - F_o, & j = i + 1, i = \overline{D + B - 2, S - 2}; \\ F_o, & j = D - 1, i = \overline{D + B - 2, (w + 1)B - 2}, w \geq K + 2; \\ F_o, & j = kB, i = \overline{D + (w - k)B - 2, D + (w - k + 1)B - 3}, \\ & k = \overline{1, w - 1}, w \leq K + 1; \\ F_o, & j = D - kB - 1, i = \overline{(w + k)B - 1, (w + k + 1)B - 2}, \\ & k = \overline{1, w - 1}, w \geq K + 2; \\ F_o, & j = 0, i = \overline{(w + K + 1)B - 1, S - 2}; \\ 1, & j = 0, i = S - 1. \end{cases}$$

A variety of types of solutions for system of equilibrium equations for probabilities of states of Markov's chain is determined by the relationships between the protocol parameters w , S , the size of block parameter B and

the total length of the tract D . Since the duration of the timeout must exceed the window width and be no shorter than the round trip delay of block segments ($S \geq D + B - 1$), there are five variants of the solutions for various areas of applications of protocol parameters.

4. Throughput analysis of selective repeat mode

The most important operating characteristic of the Protocol is its throughput, defined by the parameters of the data transmission channel and the mechanism of forward error correction, and overhead as well as the peculiarities of procedure of transmission control [1, 2]. Normalized performance of a transport connection is determined by the average number delivered to the recipient undistorted segments (including selective repeat mode [2] and a direct mechanism to restore the distorted segments) mean time between two successive arrivals of acknowledgement [3–6]. As the time between arrivals of acknowledgments distributed according to a geometric law with parameter F_o , mean time between arrivals of acknowledgements in the duration of the cycle t will be $\bar{T} = 1/F_o$. Then for the selective procedure of rejection throughput under $S \geq D + (w + 1)B - 2$ will be determined by the relation:

$$Z(w, S) = F_o \Phi \left\{ \sum_{k=1}^w Ak \sum_{i=D+kB-2}^{D+(k+1)B-3} P_i + Aw \sum_{i=D+(w+1)B-2}^{S-1} P_i \right\},$$

and for $\max\{D + B - 1, wB + 1\} \leq S \leq D + (w + 1)B - 2$ — by the relation:

$$Z(w, S) = F_o \Phi \left\{ \sum_{k=1}^L Ak \sum_{i=D+kB-2}^{D+(k+1)B-3} P_i + A(L + 1) \sum_{i=D+(L+1)B-2}^{S-1} P_i \right\}.$$

Here $\Phi = \sum_{i=A}^B C_B^i F_n^i (1 - F_n)^{B-i}$ — the probability of a direct reduction unit segments with errors (the reliability of delivery of the unit segments to the destination without retransmission), and $L = \lfloor (S - D + 2)/B \rfloor - 1$ — is the time between the duration of the timeout and the time of arrival of the acknowledgement. Hence it is easy to see that when $A = B = 1$ we get the known result [10]. For absolutely reliable reverse data path ($F_o = 1$) the throughput is determined by the relation $Z(w, S) = \frac{A\Phi}{B}$ when $w \geq K + 2$ and by the dependence $Z(w, S) = \frac{A\Phi}{D-1-(w-2)B}$ — when $w \leq K + 1$. In the case of unlimited window size ($w = \infty$) performance of a transport connection will take the form $Z(w, S) = \frac{AF_o\Phi}{[1-(1-F_o)^B](1+(B-1)F_o)}$.

For timeout of minimal duration $S = D + B - 1$ the performance of the transport connection is invariant to the size of the window $Z(w, S) =$

$\frac{AF_o\Phi}{D+B-1}$. When $w \leq K + 1$ and limitless duration of timeout ($S = \infty$) the throughput is converted to:

$$Z(w, \infty) = A\Phi P_0 \left[1 - (1 - F_o)^{wB} \right] \Bigg/ \left\{ F_o(1 - F_o)^B \left[1 - (1 - F_o)^{B(w-1)} \right] + \left[1 + (D - (w - 2)B - 2)F_o \right] \left[1 - (1 - F_o)^B \right] \right\}.$$

If $S = D + (w + 1)B - 2$, then

$$Z(w, S) = \frac{P_0 A \Phi \left\{ 1 - (1 - F_o)^{wB} \left(1 + w \left[1 - (1 - F_o)^B \right] \right) \right\}}{(1 - F_o)^{B(w-1)} \left[1 - (1 - F_o)^B \right]}.$$

From the numerical results it can be seen that the dependence of the throughput size when the window $1 \leq w < K + 1$ has a slight increase, in the area $w = K + 1$ — the sharp increase and then when $w > K + 2$ — the saturation to the limit values. The duration of the round trip delay D the throughput has an inverse (symmetric) dependency. The performance of a transport connection depends on duration timeout and has the character of a curve with saturation, and with the growth S quickly goes for extreme performance.

5. Discussion of the advantages of forward error correction mechanism

The most important task of effective use of technology of forward error correction is the choice of parameters of block length of a sequence of segments B and the number of redundant segments $B - A$ in the block for error correction, providing the fastest possible transport connection. It is obvious that the presence of excess segments in a transmitted sequence increases the probability of delivery to the recipient information segments in the group. However, this is achieved by the growth of overheads in the form of time transfer of redundant data. In this connection there is the task of searching for ranges of values of characteristics of transport connection (D, F_o, F_n), parameters of transport protocol (w, S) and the mechanism of forward error correction (A, B), providing superior management procedures with forward error correction to the classical protocol procedure without FEC. Let's do comparative analysis of the protocol procedures with and without the use of FEC mechanism. Comparison of the control procedure is performed under conditions of equal intensities of subscriber streams offered for transfer $\lambda = Aw$. We define the benefit in speed from

the use of FEC mechanism compared with the classical protocol procedure, decision feedback in the form of:

$$\Delta(w, S) = Z(D, w, S, A, B) - Z(D, Aw, S, 1, 1).$$

With a reliable reverse path data transmission $F_o = 1$ and $w \leq K + 1$ the benefit has a simple analytical form:

$$\Delta(w, S) = \frac{A\Phi}{D - 1 - B(w - 2)} - \frac{F_n}{D + 1 - Aw}.$$

In general, the comparative analysis is possible to carry out only numerically. The most significant gain is determined by the relation between the width of the window and the round-trip delay duration of the transport connection D . In general, the dependence of the benefit from this option has a positive value when the window size satisfying the constraints $1 \leq w < 2(K + 2)$. For the field of $1 \leq w < K + 1$ the subscriber stream is advantageous to configure the window width $w = 1$ and the maximum group size $B < D$, while for $w \geq K + 2$ more profitable to use of small parameters of FEC. It should also be noted that with the growth of parameters of forward error correction A and B in the area of $1 \leq w < K + 1$ the absolute value of the winnings grow, but the positive values of the winning coordinate is the reliability of the transmission of the data segments in the transport connection in this case is narrowed. Obviously, that for not fully loaded transport connection ($w \leq K + 1$) the use of FEC mechanism would be most beneficial. This is due to the fact that during idle periods of the sender waiting for acknowledgements source can download the data transmission path of the mandrel redundant segments and thereby reduce the probability of re-transmission with virtually no increase in overhead.

6. Conclusions

In this paper, a model of the process of transfer data segments in a transport connection managed by the reliable transport protocol with a mechanism of forward error correction and confirmation of data received by the receiver, in selective repeat mode. A mathematical model based on the description of the queue transmitted, but not confirmed data segments with Markov's chain with a finite number of states and discrete time. The obtained stationary distribution of different states of Markov's chain for various areas of changes in window size and the duration of the timeout. The analytical expressions for the throughput of transport connection are found. Overall throughput is largely determined by the relation between the width of the window and the round trip delay. It is shown that for the selective repeat mode, the application of the FEC mechanism is reasonable but for not fully loaded transport links ($w < K + 1$) with a large round trip delay (D).

References

1. *Fall K., Stevens R.* TCP/IP Illustrated, Volume 1: The Protocols (2nd Edition). Addison-Wesley Professional Computing Series, 2012. 1017 p.
2. *Boguslavskii L. B.* *Upravlenie potokami dannykh v setyakh EVM* (Controlling Data Flows in Computer Networks), Moscow: Energoatomizdat, 1984. 168 p.
3. *Lundqvist H., Karlsson G.* TCP with end-to-end FEC // Communications, 2004 International Zurich Seminar on. 2004. P. 152–156.
4. *Barakat Ch., Altman E.* Bandwidth tradeoff between TCP and link-level FEC // Computer Networks. 2002. No. 39. P. 133–150.
5. *Shalin R., Kesavaraja D.* Multimedia Data Transmission through TCP/IP using Hash Based FEC with AUTO-XOR Scheme // IC-TACT Journal on Communication Technology. 2012, Vol.: 03. Issue: 03. P. 604–609.
6. *Boguslavskij L.B., Gelenbe E.* Analytical models of data link control procedures in packet-switching computer networks // Automation and Remote Control. 1980. Vol. 41. No. 7. P. 1033–1042.
7. *Gelenbe E., Labetoulle J., Pujolle G.* Performance Evaluation of the HDLC Protocol // Computer Networks. 1978. Vol. 2. Issue 4/5. P. 409–415.
8. *Kokshenev V. V., Suschenko S. P.* Analysis of the asynchronous performance management procedures link transmission data // Computational technologies. 2008. Vol. 15. Special issue No. 5. P. 61–65.
9. *Kokshenev V. V., Mikheev P. A., Suschenko S. P.* Transport protocol selective acknowledgements analysis in loaded transmission data path // Vestnik TSU. Series control, computer facilities and computer science. 2013. No. 3(24). P. 78–94.
10. *Kokshenev V. V., Suschenko S. P.* Modeling sessions with Markov's chains // Theory of probability, random processes, mathematical statistics and applications: proceedings of the International scientific conference dedicated to the 80th anniversary of Professor G. A. Medvedev. Minsk on 23-26 February 2015. Minsk, RIVS, 2015. P. 311–316.
11. *Calleari C., Giordano S., Pagano M., Pepe T.* A survey of congestion control mechanisms in Linux TCP // DCCN 2013, Moscow, Russia, October 7–10, 2013. Revised Selected Papers Vishnevsky, V.; Kozyrev, D.; Larionov, A. (Eds.). 2014. P. 28–42.
12. *Kokshenev V. V., Suschenko S. P.* Analytical Model of the TCP Reno Congestion Control Procedure through a Discrete-Time Markov Chain // DCCN 2013, Moscow, Russia, October 7–10, 2013. Revised Selected Papers Vishnevsky, V.; Kozyrev, D.; Larionov, A. (Eds.). 2014. P. 124–135.
13. *Ivanovskii V. B.* Properties of Output Flows in Discrete Queueing Systems // Automation and Remote Control. 1984. Vol. 45. No. 11. Part 1. P. 1413–1419.
14. *Tannenbaum A.* Modern Operating Systems. SPb: Piter, 2002.

УДК 519.872

Суммарный объем заявок в бесконечнолинейной системе массового обслуживания с рекуррентным входящим потоком

В. А. Колбасова, Е. Ю. Лисовская, С. П. Моисеева

** Кафедра теории вероятностей и математической статистики,
Национальный исследовательский Томский государственный
университет,
пр. Ленина, д.36, Томск, Россия, 634050*

Аннотация. В данной работе рассматривается бесконечнолинейная система массового обслуживания с входящим рекуррентным потоком требований случайного объема и независимым от этого объема временем обслуживания заявок. Для такой системы методом асимптотического анализа в условии высокой интенсивности входящего потока была получена асимптотическая характеристическая функция первого порядка двумерного распределения вероятностей числа занятых приборов и суммарного объема требований, находящихся в системе.

Ключевые слова: бесконечнолинейная система массового обслуживания, требования случайного объема, рекуррентный поток, асимптотический анализ, высокоинтенсивный входящий поток.

Введение

Системы массового обслуживания (СМО) требований случайного объема имеют свое применение при моделировании процессов передачи данных, где информация передается порциями в виде сообщений случайного объема.

В работах [1–3] для наиболее простых СМО с входящим Пуассоновским потоком заявок и экспоненциальным временем обслуживания найдены характеристические функции распределения вероятностей числа заявок в системе и их суммарного объема. Однако адекватными моделями реальных процессов являются более общие модели СМО, а именно с входящим рекуррентным потоком и произвольным временем обслуживания заявок.

1. Постановка задачи

Рассмотрим систему массового обслуживания с неограниченным числом приборов, на вход которой поступает рекуррентный поток, заданный функцией распределения длин интервалов между последовательными моментами поступления заявок в систему $A(x)$. Продолжительность обслуживания заявки имеет произвольную функцию распределения, одинаковую для всех приборов $B(x)$. Предполагаем, что

каждое требование характеризуется некоторым случайным объёмом $\nu > 0$, $G(y) = P\{\nu < y\}$ – функция распределения случайного процесса ν . Объёмы различных требований независимы. По окончании обслуживания заявка покидает систему и «уносит» свой объём.

Пусть $i(t)$ – число заявок, находящихся на обслуживании в системе в момент времени t , $V(t)$ – полная сумма объёмов требований, находящихся в системе в момент времени t .

Поставим задачу нахождения характеристик двумерного случайного процесса $\{i(t), V(t)\}$. Отметим, что исследуемый процесс не является марковским. Поэтому для его исследования будем использовать метод динамического просеивания (метод просеянного потока) [4, 5].

Построим просеянный поток для рассматриваемой СМО $GI/GI/\infty$. Для этого зафиксируем некоторый момент времени T . Полагаем, что заявка входящего потока, поступившая в систему в момент времени $t < T$ с вероятностью

$$S(t) = 1 - B(T - t)$$

формирует событие просеянного потока, а с вероятностью $1 - S(t)$ эта заявка не рассматривается.

Обозначим $n(t)$ – число событий просеянного потока, наступивших до момента времени t . Тогда, если в начальный момент $t_0 < T$ система была свободна, то для момента времени T для любых m выполняется равенство

$$P\{i(t) = m\} = P\{n(t) = m\}.$$

Следует отметить, что использование метода просеянного потока позволяет более точно определить характеристики процесса $V(t)$, так как в просеянном потоке присутствуют только те заявки, которые не закончат обслуживание к моменту времени T .

2. Дифференциальное уравнение Колмогорова

Введем обозначение $P(z, n, v, t) = P\{z(t) < z, n(t) = n, V(t) < v\}$ – распределение вероятностей трехмерного марковского процесса, где $z(t)$ – остаточное время от момента t до момента наступления следующего события в исходном рекуррентном потоке, $n(t)$ – число событий просеянного потока, наступивших до момента времени t , $V(t)$ – суммарный объём требований, находящихся в просеянном потоке в момент времени t . Для этого распределения составим Δt -методом прямую систему дифференциальных уравнений Колмогорова. По формуле полной вероятности запишем равенства

$$P(z, n, v, t + \Delta t) = [P(z + \Delta t, n, v, t) - P(\Delta t, n, v, t)] + \\ + P(\Delta t, n, v, t)(1 - S(t))A(z) + \quad (1)$$

$$+S(t)A(z) \int_0^v P(\Delta t, n-1, v-y, t) dG(y) + o(\Delta t),$$

$$z > 0, n = 0, 1, 2 \dots, v > 0.$$

Из (1) получаем систему дифференциальных уравнений Колмогорова

$$\frac{\partial P(z, n, v, t)}{\partial t} = \frac{\partial P(z, n, v, t)}{\partial z} + \frac{\partial P(0, n, v, t)}{\partial z} (A(z) - 1) +$$

$$+S(t)A(z) \left[\int_0^v \frac{\partial P(0, n-1, v-y, t)}{\partial z} dG(y) - \frac{\partial P(0, n, v, t)}{\partial z} \right],$$

$$z > 0, n = 0, 1, 2 \dots, v > 0.$$

с начальным условием

$$P(z, n, v, t_0) = \begin{cases} R(z), & n = 0, v > 0, \\ 0, & \text{иначе.} \end{cases}$$

Здесь и далее $R(z)$ – стационарное распределение вероятностей значений случайного процесса $z(t)$.

Введем частичные характеристические функции вида:

$$H(z, u_1, u_2, t) = M\{exp(ju_1 n(t) + ju_2 V(t))\} = \sum_{n=0}^{\infty} e^{ju_1 n} \int_0^{\infty} e^{ju_2 v} P(z, n, v, t) dv,$$

$$z > 0, n = 0, 1, 2 \dots, v > 0.$$

Учитывая, что

$$\sum_{n=1}^{\infty} e^{ju_1 n} \int_0^{\infty} e^{ju_2 v} \int_0^v \frac{\partial P(0, n-1, v-y, t)}{\partial z} dG(y) dv =$$

$$= e^{ju_1} \sum_{n=1}^{\infty} e^{ju_1 (n-1)} \int_0^{\infty} e^{ju_2 y} e^{ju_2 (v-y)} \int_0^v \frac{\partial P(0, n-1, v-y, t)}{\partial z} dG(y) dv =$$

$$= e^{ju_1} \int_0^{\infty} e^{ju_2 y} \left[\sum_{n=1}^{\infty} e^{ju_1 (n-1)} \int_0^{\infty} e^{ju_2 (v-y)} \frac{\partial P(0, n-1, v-y, t)}{\partial z} dv \right] dG(y) =$$

$$\begin{aligned}
&= e^{ju_1} \int_0^{\infty} e^{ju_2 y} \frac{\partial H(0, u_1, u_2, t)}{\partial z} dG(y) = e^{ju_1} \frac{\partial H(0, u_1, u_2, t)}{\partial z} \int_0^{\infty} e^{ju_2 y} dG(y) = \\
&= e^{ju_1} \frac{\partial H(0, u_1, u_2, t)}{\partial z} G^*(u_2),
\end{aligned}$$

где $G^*(u_2)$ обозначено, как

$$G^*(u_2) = \int_0^{\infty} e^{ju_2 y} dG(y),$$

можно записать следующее дифференциальное уравнение:

$$\begin{aligned}
&\frac{\partial H(z, u_1, u_2, t)}{\partial t} = \frac{\partial H(z, u_1, u_2, t)}{\partial z} + \\
&+ \frac{\partial H(0, u_1, u_2, t)}{\partial z} [A(z) - 1 + S(t)A(z) (e^{ju_1} G^*(u_2) - 1)], \quad (2)
\end{aligned}$$

с начальным условием

$$H(z, u_1, u_2, t_0) = R(z). \quad (3)$$

3. Метод асимптотического анализа

Так как прямое решение уравнения (2) не представляется возможным, то для решения задачи (2) – (3) воспользуемся методом асимптотического анализа [5] в условии неограниченно растущей интенсивности входящего потока [4]. Запишем функцию распределения длин интервалов между моментами поступления заявок в систему в виде $A(Nz)$, где $N \rightarrow \infty$ – параметр высокой интенсивности потока.

Тогда, выполнив преобразования, уравнение (2) примет вид

$$\begin{aligned}
&\frac{1}{N} \frac{\partial H(z, u_1, u_2, t)}{\partial t} = \frac{\partial H(z, u_1, u_2, t)}{\partial z} + \\
&+ \frac{\partial H(0, u_1, u_2, t)}{\partial z} [A(z) - 1 + S(t)A(z) (e^{ju_1} G^*(u_2) - 1)], \quad (4)
\end{aligned}$$

с начальным условием

$$H(z, u_1, u_2, t_0) = R(z). \quad (5)$$

Асимптотический анализ первого порядка проведем в виде доказательства следующей теоремы.

Теорема. Асимптотическая характеристическая функция распределения вероятностей процесса $\{z(t), n(t), V(t)\}$ первого порядка имеет вид

$$H(z, u_1, u_2, t) = R(z) \exp \left\{ N \lambda [ju_1 + ju_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\},$$

где $\lambda = R'(0)$, a_1 – математическое ожидание случайной величины, определяемой функцией распределения объема требований $G(y)$.

Доказательство.

Выполним в выражениях (4) и (5) замены

$$\varepsilon = \frac{1}{N}, u_1 = \varepsilon w_1, u_2 = \varepsilon w_2, H(z, u_1, u_2, t) = F_1(z, w_1, w_2, t, \varepsilon). \quad (6)$$

Тогда задача (4) - (5) примет вид

$$\begin{aligned} \varepsilon \frac{\partial F_1(z, w_1, w_2, t, \varepsilon)}{\partial t} &= \frac{\partial F_1(z, w_1, w_2, t, \varepsilon)}{\partial z} + \\ &+ \frac{\partial F_1(0, w_1, w_2, t, \varepsilon)}{\partial z} [A(z) - 1 + S(t)A(z) (e^{j\varepsilon w_1} G^*(\varepsilon w_2) - 1)], \end{aligned} \quad (7)$$

с начальным условием

$$F_1(z, w_1, w_2, t_0, \varepsilon) = R(z). \quad (8)$$

Найдем асимптотическое при $\varepsilon \rightarrow 0$ решение задачи (7) - (8), то есть $F_1(z, w_1, w_2, t) = \lim_{\varepsilon \rightarrow 0} F_1(z, w_1, w_2, t, \varepsilon)$.

Этап 1. Положим в (7) $\varepsilon \rightarrow 0$, получим

$$\frac{\partial F_1(z, w_1, w_2, t)}{\partial z} + \frac{\partial F_1(0, w_1, w_2, t)}{\partial z} (A(z) - 1) = 0.$$

Можем сделать вывод, что $F_1(z, w_1, w_2, t)$ может быть представлена в виде

$$F_1(w_1, w_2, t) = R(z) \Phi_1(w_1, w_2, t), \quad (9)$$

где $\Phi_1(w_1, w_2, t)$ – некоторая скалярная функция, в силу (8), удовлетворяющая условию $\Phi_1(w_1, w_2, t_0) = 1$.

Этап 2. Выполним в (7) предельный переход при $z \rightarrow \infty$. Получим

$$\varepsilon \frac{\partial F_1(\infty, w_1, w_2, t, \varepsilon)}{\partial t} = \frac{\partial F_1(0, w_1, w_2, t, \varepsilon)}{\partial z} S(t) (e^{j\varepsilon w_1} G^*(\varepsilon w_2) - 1).$$

Подставим сюда выражение (9), воспользуемся разложениями

$$e^{j\varepsilon w_1} = 1 + j\varepsilon w_1 + O(\varepsilon^2), e^{j\varepsilon w_2} = 1 + j\varepsilon w_2 + O(\varepsilon^2),$$

поделим обе части на ε и произведем предельный переход при $\varepsilon \rightarrow 0$. С учетом того, что $R'(0) = \lambda$ [4], получим дифференциальное уравнение относительно функции $\Phi_1(w_1, w_2, t)$

$$\frac{\partial \Phi_1(w_1, w_2, t)}{\partial t} = \Phi_1(w_1, w_2, t) [\lambda S(t)(jw_1 + jw_2 a_1)], \quad (10)$$

здесь и далее $a_1 = \int_0^\infty y dG(y)$ – математическое ожидание случайной величины, определяемой функцией распределения объема требований $G(y)$. Решение (10) с учетом начального условия дает

$$\Phi_1(w_1, w_2, t) = \exp \left\{ \lambda(jw_1 + jw_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}.$$

Подставляя данное выражение в (9), получаем

$$F_1(w_1, w_2, t) = R(z) \exp \left\{ \lambda(jw_1 + jw_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}.$$

В силу замен (6) можно записать асимптотическое, при $\varepsilon \rightarrow 0$, приближенное равенство:

$$\begin{aligned} H(z, u_1, u_2, t) &= F_1(z, w_1, w_2, t, \varepsilon) \approx F_1(z, w_1, w_2, t) = R(z) \Phi_1(w_1, w_2, t) = \\ &= R(z) \exp \left\{ \lambda [jw_1 + jw_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\} = \\ &= R(z) \exp \left\{ N\lambda [ju_1 + ju_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\}. \end{aligned}$$

Теорема доказана.

Функция $H(u_1, u_2, t) = \lim_{z \rightarrow \infty} H(z, u_1, u_2, t)$ есть характеристическая функция для процесса $\{n(t), V(t)\}$ – числа событий, наступивших в просеянном потоке к моменту времени t , и суммарного объема требований в просеянном потоке к моменту времени t .

Следствие 1. *Полагая $t = T$, $t_0 = -\infty$, для характеристической функции процесса $\{i(t), V(t)\}$ в стационарном режиме получим*

$$H(u_1, u_2) = \exp\{N\lambda b_1 [ju_1 + ju_2 a_1]\},$$

здесь и далее

$$b_1 = \int_{-\infty}^T S(\tau) d\tau = \int_0^{\infty} (1 - B(\tau)) d\tau$$

определяет математическое ожидание случайной величины с функцией распределения $B(x)$.

Следствие 2. *Асимптотическая характеристическая функция суммарного объема требований в системе первого порядка в стационарном режиме имеет вид:*

$$H(u) = \exp\{juN\lambda b_1 a_1\}.$$

Заключение

Для бесконечнолинейной системы массового обслуживания требований случайного объема была получена асимптотическая характеристическая функция первого порядка двумерного распределения вероятностей числа занятых приборов и суммарного объема требований, находящихся в системе.

Литература

1. Колбасова В.А., Лисовская Е.Ю. Исследование суммарного объема требований в СМО вида GI/M/∞ // Научное творчество молодежи. Математика. Информатика : материалы XX Всероссийской научно-практической конференции (28-29 апреля 2016 г.) / сост. Ю.А. Намкина. – Томск : Изд-во Том. ун-та, 2016. – Ч.1. – с. 97-101.
2. Лисовская Е.Ю. Характеристическая функция распределения вероятностей суммарного объема заявок в системе M/GI/∞ // Материалы 54-й Международной научной студенческой конференции МНСК-2016: Математика / Новосиб. гос. ун-т. Новосибирск, 2016. – с. 108.
3. Лисовская Е.Ю., Моисеева С.П. Исследование суммарного объема требований в бесконечнолинейной системе массового обслуживания вида M/GI/∞ // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем : материалы Всероссийской конференции с международным участием. Москва, РУДН, 18-22 апреля 2016 г. – Москва : РУДН, 2016. – с. 28-30.

4. *Моисеев А.Н., Назаров А.А.* Бесконечнолинейные системы и сети массового обслуживания. —Томск: Изд-во. НТЛ, 2015. —240с.
5. *Назаров А.А., Моисеева С.П.* Метод асимптотического анализа в теории массового обслуживания. —Томск: Изд-во. НТЛ, 2015. —112с.

UDC 519.872

Total capacity of customers in infinite-server queueing system with stationary renewal arrivalse

V. Kolbasova, E. Lisovskaya, S. Moiseeva

*Department of probability theory and mathematical statistics,
Tomsk State University
Russia, 634050, Tomsk, 36 Lenin Pr.*

In the paper, the infinite-server queueing system with a random capacity of customers is considered. In this system, the total capacity of customers is analysed by means of the asymptotic analysis method with high-rate stationary renewal process arrivals.

Keywords: infinite-server queueing system, customer with random capacity, stationary renewal process.

УДК 621.391

Программное обеспечение для тестирования контроллеров программно-конфигурируемой сети

А. О. Колечкин, А. Г. Владыко

*Кафедра сетей связи и передачи данных,
Санкт-Петербургский университет телекоммуникаций им. проф. М.А.
Бонч-Бруевича,
пр. Большевиков, д.22, Санкт-Петербург, Россия, 193232*

Аннотация.

Статья посвящена исследованию существующего программного обеспечения для тестирования контроллеров программно-конфигурируемых сетей, находящегося в свободном доступе. В качестве метода исследования применяется натурный эксперимент. В работе оценивается функциональность исследуемых средств тестирования, а так же их удобство использования. В статье приведены результаты оценки характеристик утилит для тестирования контроллеров программно-конфигурируемых сетей находящихся в свободном доступе, таких как NSProbe, Sbench, CtlTest. Полученные результаты позволят выбрать подходящее программное обеспечение для тестирования ПКС контроллеров при планировании сетей с различными требованиями, а так же помогут определить недостатки и вектор дальнейшего развития существующего ПО и разработать требования к программному обеспечению для тестирования SDN контроллеров, которое будет разрабатываться впоследствии.

Ключевые слова: программно-конфигурируемая сеть, коммутатор, контроллер, OpenFlow, тестирование, ПКС.

1. Введение

В настоящее время остро стоит проблема перегруженности сетей связи NGN. Учитывая тот факт, что к 2020 году к сетям общего пользования будут подключено более 50 миллиардов устройств Интернета вещей, Промышленного интернета, есть риск, что сети не справятся с трафиком, генерируемым таким количеством узлов [1–5]. Также появится сложность в управлении и конфигурировании сетей такого масштаба.

Данные недостатки современных сетей позволит исправить технология SDN (Software Defined Networking) [6]. SDN предполагает логически централизованное управление сетью за счет программирования. Программируемые сети позволять гибко распределять и обрабатывать большие потоки трафика, устранять узкие места и резервировать ресурсы сети, предотвращая перегрузку [7]. Централизованный метод управления делает контроль и настройку большого количества устройств проще [8].

Концепция SDN предполагает:

- Разделение процессов передачи данных и управления данными;

- Логически централизованный уровень управления данными;
- Виртуализация физических ресурсов сети;
- Единый и унифицированный интерфейс (OpenFlow) между плоскостью управления и плоскостью передачи данных (не зависит от производителя).

Архитектура ПКС состоит из трех уровней:

- Уровень сетевых приложений: на данном уровне реализуются различные функции управления сетью: управление потоками данных в сети, управление безопасностью, мониторинг трафика, управление QoS, управления политиками и так далее [9];

- Уровень управления: на уровне управления отслеживается и поддерживается глобальное представление сети (топология сети). Также на этом уровне реализуется программный интерфейс (API) для сетевых приложений [10, 11];

- Уровень инфраструктуры сети: включает в себя сетевые устройства ПКС сети (коммутаторы OpenFlow) и каналы передачи данных [12].

С одной стороны, такой подход дает большую гибкость в управлении сетью и существенно упрощает администрирование. С другой стороны, появление новых технологий управления сетью сопровождается отсутствием средств для их эталонного тестирования. В результате потенциальные потребители таких технологий не всегда могут оценить заявленные разработчиками технологические возможности.

Таким образом, целью данного исследования является определение характеристик существующих средств для тестирования главного элемента SDN сети – контроллера, выявление их недостатков и определение пути их последующего развития.

2. Характеристики контроллера SDN

Контроллер является ключевым элементом ПКС сети, он выступает в роли «мозга» всей сети. Производительность и возможности сети напрямую связаны с характеристиками контроллера [13]. Сам контроллер представляет собой сетевую операционную систему, установленную на выделенном физическом сервере.

К основным функциям контроллера относятся:

- Управление устройствами сети;
- Управление топологией (построение топологии сети, обработка добавления/удаления новых элементов сети).
- Управление приложениями.
- Управление доступными ресурсами сервера (потоками, ядрами).

К характеристикам контроллера относятся:

Производительность:

- Пропускная способность - количество запросов от коммутаторов, обрабатываемых контроллером в секунду - [потоки/секунду];

- Задержка - время, затрачиваемое контроллером на обработку одного запроса - [миллисекунды].

Масштабируемость:

- Изменение показателей производительности при увеличении числа соединений с коммутаторами;

- Изменение показателей производительности при увеличении числа конечных узлов в сети;

- Изменение показателей производительности при увеличении числа ядер процессора.

Функциональность:

- Корректность обработки контроллером сообщений протокола OpenFlow от коммутаторов.

Безопасность:

- Устойчивость контроллера ПКС к некорректно сформированным сообщениям протокола OpenFlow.

Ресурсоемкость:

- Загрузка ядер процессора;

- Использование физической памяти.

Надежность:

- Количество отказов за время тестирования;

- Время безотказной работы при заданном профиле нагрузок.

3. Программное обеспечение для тестирования контроллеров SDN

3.1. Cbench

Утилита Cbench (controller benchmarker) на сегодняшний день является одним из немногих доступных инструментов для проведения тестирования производительности контроллеров. Cbench входит в состав OpenFlow платформы тестирования OFlops [14]. Ввод параметров тестирования и вывод результатов производится с помощью интерфейса командной строки.

Измеряемые характеристики. Текущие показатели, которые измеряются с помощью Cbench:

1. Максимальное, среднее, минимальное количество ответов контроллера в секунду на сообщения packet_in.
2. Максимальная, средняя, минимальная задержка между прибытием пакета и соответствующим сообщением packet_in.

Cbench поддерживает два режима работы:

- Режим измерения пропускной способности (Throughput mode). В режиме пропускной способности каждый коммутатор отправляет сообщения контроллеру, не дожидаясь от него ответа, до тех пор, пока есть место в буфере отправки. Этот режим позволяет оценить количество запросов, которое контроллер может обрабатывать за единицу времени при максимальной нагрузке.

- Режим измерения задержки (Latency mode). В режиме задержки каждый коммутатор отправляет одно сообщение контроллеру и дожидается от него ответа, прежде чем отправить следующее, при этом отправление сообщений другими коммутаторами не блокируется.

По результатам работы каждого теста определяется суммарное количество потоков, установленных на всех коммутаторах, и среднее количество потоков, устанавливаемых контроллером в секунду (flow/s).

В результате исследования данного ПО было определено, что Sbench позволяет проводить только тестирование производительности и масштабируемости SDN контроллера, не охватывая при этом остальные характеристики контроллера. Настройка параметров тестирования выполняется с помощью определенных команд интерфейса командной строки и позволяет довольно гибко задавать конфигурацию проводимого теста, но требует времени для освоения пользователем. Анализ результатов также представляется довольно сложным для неопытного пользователя, ввиду вывода их в виде массива строк в интерфейс командной строки.

3.2. HCProbe

Программное средство Hcprobe разработки Центра прикладных исследований компьютерных сетей. Используемый язык программирования Haskell [15].

Можно выделить следующие особенности HCProbe:

1. HCProbe позволяет генерировать произвольные сообщения OpenFlow, изменять значение любого поля заголовка OpenFlow сообщения или его содержимое.
2. HCProbe реализован на языке программирования высокого уровня Haskell, что упрощает сопровождение и расширение функциональности данного средства.
3. HCProbe предоставляет гибкий API для разработки тестовых сценариев на встроенном языке EDSL.

Архитектура HCProbe, согласно Рисунку 1, содержит следующие основные модули:

- Модуль Network datagram. Библиотека, предоставляющая интерфейс для генерации сообщений всех протоколов сетевого стека (кадры Ethernet, пакеты IP и ARP, пакеты TCP и UDP). Интерфейс предоставляет возможность задавать произвольные значения полей заголовков и генерировать произвольное содержимое пакета.

- Модуль OpenFlow. Библиотека для генерации и обработки OpenFlow сообщений.

- Модуль Configuration. Библиотека, предоставляющая интерфейс для задания параметров теста при помощи опций командной строки или конфигурационного файла.

- Модуль FakeSwitch. Модуль, эмулирующий работу коммутаторов, которые устанавливают соединения с контроллером. Данный модуль

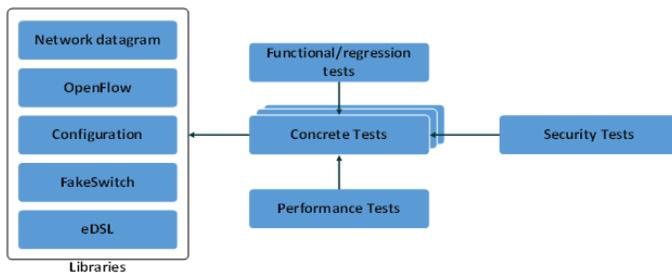


Рис. 1. Архитектура HCPProbe

предоставляет интерфейс обработки и отправки сообщений контроллеру.

- Модуль eDSL. Модуль предоставляет упрощенный интерфейс для написания тестовых сценариев.

Основные настраиваемые параметры в HCPProbe могут быть заданы при запуске при помощи опций командной строки или в конфигурационном файле.

В результате исследования данного ПО было определено, что HCPProbe позволяет проводить тестирование более обширного списка характеристик чем Sbench, а именно производительность, безопасность, масштабируемость и функциональность, но не охватывает такие характеристики контроллера, как надежность и ресурсоемкость. Настройка параметров тестирования с использованием предоставляемых сценариев выполняется с помощью определенных команд интерфейса командной строки и позволяет довольно гибко задавать конфигурацию проводимого теста, но создание собственных сценариев требует от пользователя знания языка EDSL и Haskell. Анализ результатов, как и в предыдущем ПО, представляется довольно сложным для неопытного пользователя, ввиду вывода их в виде массива строк в интерфейс командной строки.

3.3. CtlTest

Ctltest - это собрание сценариев для тестирования производительности SDN контроллеров поддерживающих протокол OpenFlow. Сценарии используют Sbench для измерения пропускной способности и задержки. Это собрание включает сценарии для семи популярных OpenFlow контроллеров.

Также предоставляется возможность для добавление других контроллеров для тестирования [16]. Сценарии устанавливают и запускают контроллеры на одном сервере и, затем, по SSH запускают Sbench

на другом. Cbench отправляет сообщения формата PacketIn на контроллер и регистрирует ответы контроллера(сообщения FlowMod или PacketOut).

Сценарии запускают несколько тестов, которые помогают анализировать взаимосвязь между производительностью контроллера и количеством доступных ядер ЦП, количества коммутаторов и хостов. Также предоставляется инструмент для построения графиков этой зависимости.

Проект содержит следующие сценарии:

- `Install_contr.sh`: установка и запуск популярных OpenFlow контроллеров (NOX, POX, Floodlight, Trema, Beacon, MuL, Maestro, Ryu);
- `benchmark_throughput.sh`: измерение пропускной способности тестируемых контроллеров с использованием утилиты Cbench;
- `benchmark_latency.sh`: измерение задержки тестируемых контроллеров с использованием утилиты Cbench;
- `stat.sh`: сбор статистики использования CPU и физической памяти во время запуска контроллеров;
- `plotter.py`: построение графиков результатов тестирования с использованием gnuplot.

В результате исследования было определено, что CtlTest позволяет проводить тестирование производительности, масштабируемость и ресурсоемкости контроллеров, но не охватывает такие характеристики контроллера, как надежность, безопасность и функциональность. Помимо этого CtlTest предоставляет возможность для визуализации полученных результатов в виде линейных диаграмм, что значительно упрощает их анализ.

4. Заключение

В ходе исследований существующих решений, находящихся в свободном доступе, для тестирования контроллеров программно-конфигурируемых сетей были выявлены следующие недостатки:

- Отсутствие средства для полного тестирования всех характеристик контроллера;
- Отсутствие графического интерфейса.

Таким образом, по-прежнему актуальным является проблема разработки платформы или средства для комплексного анализа характеристик контроллеров. На основе рассмотренных программных решений предполагается разработать собственное средство, главными требованиями к которому будут расширенный функционал, позволяющий тонкую настройку параметров тестирования, охват всех параметров тестируемого контроллера SDN, а также наличие интуитивно понятного графического интерфейса пользователя, для ускорения работы с программным обеспечением и упрощения анализа получаемых результатов в ходе тестирования. Данное программное обеспечение предлагается использовать для тестирования узлов программно-конфигурируемого сегмента летающей сенсорной сети.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта N15-07-09431a "Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей".

Литература

1. Киричек Р. В., Парамонов А. И., Прокопьев А. В., Кучерявый А. Е. Эволюция исследований в области беспроводных сенсорных сетей // Информационные технологии и телекоммуникации. — 2014. — № 4 (8). — С. 29-41.
2. А.Е.Кучерявый Интернет Вещей // Электросвязь. — 2013. — № 1. с.21-24.
3. А.Е.Кучерявый, А.Е.Прокопьев, Е.А.Кучерявый Самоорганизующиеся сети. — СПб.: Любавич, 2011. — 312 с.
4. Kirichek R., Koucheryavy A. Internet of Things Laboratory Test Bed // Lecture Notes in Electrical Engineering - Heidelberg: Springer, 2016. Vol. 348. — PP. 485-494.
5. Kirichek R., Golubeva M., Kulik V., Koucheryavy A. The home network traffic models investigation // 18th international conference on advanced communication technology (ICACT) — Phoenix Park, Korea: IEEE, 2016. — PP. 97-100
6. Kirichek R., Vladyko A., Zakharov M., Koucheryavy A. Model networks for internet of things and sdn // 18th international conference on advanced communication technology (ICACT) — Phoenix Park, Korea: IEEE, 2016. — PP. 76-79
7. S. Dotcenko, A. Vladyko, and I. Letenko, "A fuzzy logic-based information security management for software-defined networks," in Advanced Communication Technology (ICACT), 2014 16th International Conference on, Feb 2014.
8. Kreutz D. Software-Defined Networking: A Comprehensive Survey / D. Kreutz, F. M. V. Ramos, P. Esteves Verissimo, C. Esteve Rothenberg, S. Azodolmolky, S. Uhlig // Proceedings of the IEEE, 2015, Vol. 103, Iss. 1. - PP. 14-76.
9. Wenfeng Xia A Survey on Software-Defined Networking / Xia Wenfeng, Wen Yonggang, Chuan Heng Foh, D. Niyato, Xie Haiyong // IEEE Communications Surveys & Tutorials, 2015, Vol. 17, Iss. 1. - PP. 27-51.
10. Open Networking Foundation. "Software-Defined Networking: The New Norm for Networks" // ONF White Paper // April 13, 2012.
11. Киричек Р. В., Владыко А. Г., Захаров М. В., Кучерявый А. Е. Модельные сети для интернета вещей и программируемых сетей // Информационные технологии и телекоммуникации. — 2015. — № 3 (11). — С. 17-26.

12. Васильев А.Б., Тарасов Д.В., Андреев Д.В., Кучерявый А.Е. Тестирование сетей связи следующего поколения. М.:ФГУП ЦНИИС, 2008 г. – 144 с.
13. Recommendation ITU-T Q.3900 (09/2006). Methods of testing and model network architecture for NGN technical means testing as applied to public telecommunication networks. ITU, 2006.
14. OpenFlow Benchmarking [Электронный ресурс] – Режим доступа: <http://www.openflow.org/wk/index.php/Oflops>
15. Описание утилиты HCprobe [Электронный ресурс] – Режим доступа: arccn.github.io/hcprobe/
16. Ctltest [Электронный ресурс] – Режим доступа: <http://arccn.github.io/ctltest/>

UDC 621.391

Software for testing of controllers in software-defined networks

A. O. Kolechkin, A. G. Vladyko

*Department communications networks and data transmission,
Saint Petersburg State University of Telecommunications. prof.
Bonch-Bruевич,
pr.Bolshevikov d.22, Russia, St. Petersburg, 193232*

The article is devoted to study the existing software for test controllers, software defined networking, which is free to access. As a method of study used the natural experiment. The paper evaluated the functionality of the investigated testing tools and their usability. The article presents the evaluation of tools for testing controllers software defined networking which is free to access, such as HCProbe, Cbench, CtlTest. The obtained results allow to choose the right software for testing SDN controllers in the planning of networks with different requirements, as well as help to identify gaps and further development of existing software and to develop software requirements for testing SDN controllers, which will be developed subsequently.

Keywords: software defined network, controller, OpenFlow, testing, SDN.

UDC 004.056.53

The Fault-tolerant Structure of Multilevel Secure Access to the Resources of the Public Network

V. S. Kolomoitcev, V. A. Bogatyrev

*Department of Computation Technologies
ITMO University
Kronverksky Pr. 49, St. Petersburg, 197101, Russia*

Abstract. The paper presents the evaluation of the effectiveness of the structural organization of the system of multi-level secure access to external network resources. We conducted a comparative analysis and optimization of the pattern of access 'Direct connection', with of its various forms of implementation during the organization of a secure connection of end-node internal network to the resources located in the external network. The study was conducted on the basis that each security element is included in the pattern of the secure access is able to detect and eliminate the threats of the other elements of the system of protection. Pattern of access 'Direct connection' in a general form has four variants of construction, differing from each other by mutual arrangement of the key elements: firewall with packet-filtering, firewall with adaptive detailed packet inspection and the router. It was a mathematical model to calculate the reliability of the ways of construction of the pattern of access. It is shown that the most reliable way of construction of pattern of access is one that includes a single group of routers for the entire system. Ways are not very different from each other reliability value that include two groups of routers on the overall system.

Keywords: firewalls; corporate networks; information security; fault tolerance; access pattern; reliability; networking.

1. Introduction

Modern computer networks, both corporate and public, have a complicated structure. In such networks, there are some very serious problems of information security. They may be at risk of unauthorized access, denial-of-service (DoS) nodes, the loss of transmitted information, as well as threats of violations of privacy that could lead to significant economic and other losses [1].

Threats can be both external as a result of remote network attacks, and internal by various stowing software or hardware. To eliminate the challenges of information security, some measures can be taken and means of information security, located on various levels of the network used.

The principles of organization of a secure connection of the corporate network to public network are among the most important elements for ensuring information security. They have a significant impact on the safety and reliability of the network. However, it is worth remembering that the most effective security techniques usually imply some significant costs.

In this study, we investigate possibilities of the pattern for the organization secure access to external network resources, taking into account

the requirements set out in the guidance documents (legislative and legal documents) on information security. The study is aimed at a choice of rational options for creating protection system, with ensuring high levels of reliability [2, 3].

2. Object and objectives of the study

The pattern which is regarded in this paper is focused on improving the level of protection devices on the network. The key challenge of the pattern is to organize secure access to poorly protected and / or uncontrolled portions of the network. This pattern allows reducing the threat of DDoS-attacks, unauthorized access to a network node, listening to the information channel and penetration of malicious software [4].

The pattern under consideration is based on a standard network access pattern to the resources of the external network: the node 'Internal (local) network' Routers 'External network (the Internet)'. This approach minimizes the degree of possible reorganization of existing corporate network.

In the standard pattern of access the end-node of the corporate network to nodes of the external network the protection of this node is based on a built-in means: antivirus protection (AV), a standard firewall (FW), and possible means of protection against unauthorized access. In the standard pattern at the entrance to the network have a router.

The measures used in this pattern, leading to the fact that almost all of the work to eliminate threats from the external network rests on the end-node. For mission-critical systems of mentioned above means of protection are not enough. Therefore, we should use the pattern that ensures a comprehensive information security. In the role of such a pattern can be used the pattern Direct connection [4].

As a result, we are suggested to consider the possibilities of the pattern Direct connection. The aim of this study is to explore the possibilities of various physical interpretations of the pattern Direct connection in terms of ensuring high reliability of the system.

3. The basic version pattern of access 'Direct connection'

Using the pattern Direct connection involves minimal changes in the architecture of the corporate network, as well as minimal additional financial costs to implement it. The structure of the pattern "Direct connection" is presented in Fig. 1.

In this pattern, at the entrance to the internal network (just after the router) is set firewall with packet-filtering (FW-1) to eliminate spam, reduce the total load of the channel within the network, as well as reduce the risk of DDoS-attacks. Firewall with adaptive detailed packet inspection (FW-2) installed behind a FW-1 for a deeper analysis of the contents of packages. Often available as part of pattern routers can carry functional FW-1 [5]. However, the computing system architecture is not known in

advance (which will be implemented by the pattern of access). The router and the firewall will take account of what how the different elements of the system, where the router (as an element) is intended primarily for communication with parts of the system together, and the FW-1 – filter traffic in a computer system. Otherwise, routers (for communicating between a firewall in this pattern) may be combined into a single cluster and replaced with a cluster of firewall packet filtering.

Once the data have passed the FW-2, they are (potentially 'clean-data') must be received to the desired end-node. On end-node there are local antivirus (AV) (with personal firewall) installed [5], as well as some systems of protection against unauthorized access (UAA), and some secure data storage in order to reduce the negative effects of potential insider attacks. In this pattern of access, the channel data is to be protected, thereby reducing or even prevents the possibility of influence an intruder on data flowing in the channel.

To improve the overall network protection from DDoS-attacks, data loss or destruction and other threats, mission critical nodes should be reserved, and for the data stored on them, backups are created.

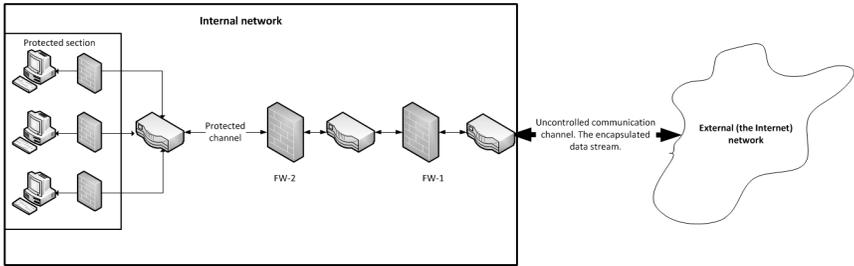


Figure 1. The pattern Direct connection

4. Ways of construction the network infrastructure pattern of access 'Direct connection'

For qualitative and uninterrupted operation of the network, you must do a backup of system components. Network architecture of the pattern 'Direct Connection' has three main components: firewall with packet-filtering, firewall with adaptive detailed packet inspection and routers that connect all the elements of pattern together. In this pattern there are four possible ways for the construction of pattern: with three, two or one groups of routers on the entire system.

Possible ways of constructing a network infrastructure pattern Direct Connection presented in Fig. 2

5. Estimation of reliability of the system

Estimation of possible ways pattern Direct connection requires a search for multiplicity of redundant nodes in each group. Required reach the highest possible level of reliability of the system, given the limitations imposed on the cost of implementation [6–8].

Reliability of the system consisting of several nodes is defined as the product of the reliability of each of the elements (groups) of the system. Reliability of the proposed patterns is equal: $P_1 = P_{01} \cdot P_{m1} \cdot P_{02} \cdot P_{m2} \cdot P_{03}$; $P_{2-3} = P_{01} \cdot P_{m1} \cdot P_{02} \cdot P_{m2}$; $P_4 = P_{01} \cdot P_{m1} \cdot P_{m2}$.

where $P_{m1} = (1 - (1 - r_1)^{n_1})$, $P_{m2} = (1 - (1 - r_2)^{n_2})$. Assuming that the routers in each group are the same: $P_{0i} = (1 - (1 - r_0)^{n_{0i}})$. Here

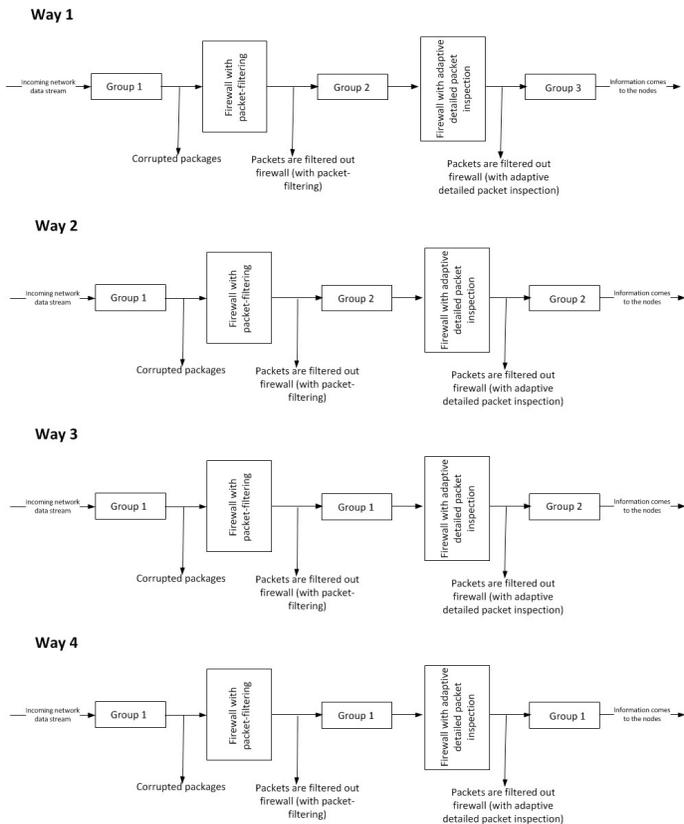


Figure 2. Variations network architecture pattern Direct Connection

$r_j = e^{-\lambda_j t}$ and $\lambda_0, \lambda_1, \lambda_2$ - failure rate of routers, FW-1 and FW-2; n_{0i} - the number of routers in the i -th group; n_1 - the number of FW-1; n_2 - the number of FW-2.

Costs for the implementation of the ways of construction of the scheme are shown in (1) and defined as:

$$C_{1-4} = c_0 \cdot \sum_i n_{0i} + c_1 \cdot n_1 + c_2 \cdot n_2. \quad (1)$$

Here c_0, c_1, c_2 - the cost of router, FW-1, FW-2.

Estimation of protection systems includes finding the distribution of each type of node that provides maximum reliability of the entire system considering the limitation of the cost of implementation: $C_1 \leq C, C_2 \leq C, \dots, C_4 \leq C$; and compliance steady state conditions of service [9–13].

After passing through the router the input flow is filtered and, thus, the density of the flow on the FW-1 is lower than the router. The same happens with the input flow received at the FW-2. After passing through the FW-1 a certain proportion of the input flow is filtered and to the FW-2 is received smaller input flow. As a result of limitations imposed on the capacity of each element of the system will be equal to:

$L_1 < 1, L_2 < 2, d_2 \cdot R/n_{02} < 1, L_4 < 1, d_4 \cdot R/n_{03} < 1, d_5 \cdot R/n_{03} < 1$
for Way 1.

$L_1 < 1, L_2 < 2, d_2 \cdot R/n_{02} < 1, L_4 < 1, d_4 \cdot R/n_{02} < 1, d_5 \cdot R/n_{02} < 1$
for Way 2.

$L_1 < 1, L_2 < 2, d_2 \cdot R/n_{01} < 1, L_4 < 1, d_4 \cdot R/n_{02} < 1, d_5 \cdot R/n_{02} < 1$
for Way 3.

$L_1 < 1, L_2 < 2, d_2 \cdot R/n_{01} < 1, L_4 < 1, d_4 \cdot R/n_{01} < 1, d_5 \cdot R/n_{01} < 1$
for Way 4.

Here $L_1 = R/n_{01}$; $L_2 = d_1 \cdot \lambda \cdot V_1/n_1$; $L_4 = d_3 \cdot \lambda \cdot V_2/n_2$; $R = \lambda \cdot V_0$ where d_i - the proportion of the filtered input flow of previously placed node; V_0, V_1, V_2 - average service time of request in routers, FW-1 and FW-2; λ - the arrival rate of requests; n_{0i} - the number of routers in the i -th group; n_1 - the number of FW-1; n_2 - the number of FW-2.

The proportion of the filtered input flow of previously placed node can be obtained by the equations (2), (3), (4), (5), (6).

After first router:

$$d_1 = (1 - A_0 \cdot p_0). \quad (2)$$

After first router and FW-1:

$$d_2 = 1 - (p_1 \cdot (p_1 \cdot (A_1 - l_{10}) + p_0 \cdot (A_0 - l_{10}) + l_{10} \cdot (1 - \bar{p}_0 \cdot p_1))). \quad (3)$$

After first router, FW-1 and second router:

$$d_3 = 1 - (p_1 \cdot (p_1 \cdot (A_1 - l_{10}) + (1 - \bar{p}_0^2) \cdot (A_0 - l_{10}) + l_{10} \cdot (1 - \bar{p}_0^2 \cdot p_1))). \quad (4)$$

After first router, FW-1, second router and FW-2:

$$d_4 = 1 - (p_1 \cdot M_{e1} + (1 - \bar{p}_0^2) \cdot (R_{emp}) + (p_2 \cdot M_{e2}) + (l_{10} - l_{00}) \cdot (1 - \bar{p}_0 \cdot \bar{p}_1 \cdot \bar{p}_2)) + (l_{20} - l_{00}) \cdot (1 - \bar{p}_0^2 \cdot \bar{p}_2) + (l_{21} - l_{00}) \cdot (1 - \bar{p}_1 \cdot \bar{p}_2) + l_{00} \cdot (1 - \bar{p}_0^2 \cdot \bar{p}_1 \cdot \bar{p}_2). \quad (5)$$

After first router, FW-1, second router, FW-2 and third router:

$$d_5 = 1 - (p_1 \cdot M_{e1} + (1 - \bar{p}_0^3) \cdot (R_{emp}) + (p_2 \cdot M_{e2}) + (l_{10} - l_{00}) \cdot (1 - \bar{p}_0 \cdot \bar{p}_1 \cdot \bar{p}_2)) + (l_{20} - l_{00}) \cdot (1 - \bar{p}_0^3 \cdot \bar{p}_2) + (l_{21} - l_{00}) \cdot (1 - \bar{p}_1 \cdot \bar{p}_2) + l_{00} \cdot (1 - \bar{p}_0^3 \cdot \bar{p}_1 \cdot \bar{p}_2). \quad (6)$$

where $R_{emp} = (A_0 - l_{20} - l_{10} + l_{00})$; $M_{e1} = (A_1 - l_{21} - l_{10} + l_{00})$; $M_{e2} = (A_2 - l_{21} - l_{20} + l_{00})$. At the same time A_0, A_1, A_2 - respectively, the proportion of threats (errors) in the input stream, the router detected with a probability p_0 ; FW-1 with a probability p_1 ; FW-2 with a probability p_2 .

Results of reliability calculation, depending on the constraints imposed on the system throughput determined by a known the arrival rate λ , when $r_0 = 0.85$, $r_1 = 0.9$; $V_0 = 0.025$ sec, $V_1 = 0.04$ sec, $V_2 = 0.075$ sec, $c_0 = 10$ cu, $c_1 = 20$ cu, $c_2 = 35$ cu, $C = 500$ cu, $p_0 = 0.85$, $p_1 = p_2 = 0.899$; $A_0 = 0.07$, $A_1 = 0.15$, $A_2 = 0.26$, $l_{00} = 0.04$, $l_{10} = 0.04$, $l_{20} = 0.06$, $l_{21} = 0.12$ are shown in Fig 3.

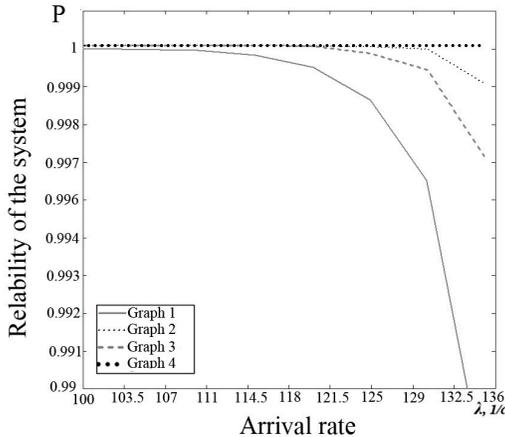


Figure 3. Reliability pattern of access at a fixed time of operation of the computer system based on the arrival rate of requests

Sequence numbers of curves in the graphs have a direct accordance with sequence numbers of their defining formulas, namely the graphs 1-4 – ways of constructing pattern "Direct connection".

As shown in Fig 3, for small values of the arrival rate, the reliability of each of the ways of the pattern Direct Connection is approximately equal. However, with increasing the arrival rate is detected, the fourth way of pattern is more reliable than other ways, and the first - the least reliable. Also from Fig 3 shows that if you want to choose one of two ways contain two groups of routers, it is a bit more reliable to take the Way 2 than Way 3.

As a result, we can conclude that if you want to use the most reliable way of the pattern, it is best to choose the Way 4. In the case where raises the question of maximizing the system reliability and to select one of two options - the Way 2 or Way 3 then there is no much difference which of them to use.

6. Conclusion

The paper analyzes possibilities pattern of access Direct Connection that allow you to organize a secure connection the end-node internal network to resources located in the external network. The study identified the advantages and disadvantages of the pattern Direct connection, depending on the way of its structure.

It has been shown that one of the ways of the pattern Direct connection (using a single group router to the entire pattern of access), has a higher level of reliability than others. We also show that the second and third variants of pattern are almost identical to each other in terms of reliability.

Thus, when implementing the pattern Direct connection is best to use a way of its structure using a common pool of routers for the entire system.

In this study, published work on finding the fastest (minimum average residence time of the request in the system) pattern configuration [4,5,14].

Acknowledgments

The work is partially supported by Government of St. Petersburg grant.

References

1. T.I.Aliev, M.I.Rebezova, A.A.Russ Statistical Methods for Monitoring Travel Agencies Automatic Control and Computer Sciences, Vol. 49, No. 6, 2015, pp. 321–327.
2. V.A.Bogatyrev, S.V.Bogatyrev, I.Y.Golubev Optimization and the Process of Task Distribution between Computer System Clusters Automatic Control and Computer Sciences, No. 3, 2012, pp.103-111.

3. Arustamov S.A., Bogatyrev V.A., Polyakov V.I. Back Up Data Transmission in Real-Time Duplicated Computer Systems Proceedings of the First International Scientific Conference Intelligent Information Technologies for Industry (IITI16), Advances in Intelligent Systems and Computing 451 Springer International Publishing Switzerland 2016, pp 103-109.
4. V.S.Kolomoitcev A comparative analysis of approaches to organizing of secure connection of the corporate network nodes to the public network Cybernetics and Programming, No. 2, 2015, pp. 46-58.
5. V.S.Kolomoitcev Choice of option for implementation of the multilevel secure access to the external network Scientific and Technical Journal of Information Technologies, Mechanics and Optics, vol. 16, no. 1, 2016, pp. 115-121.
6. V.A.Bogatyrev, A.V.Bogatyrev Functional Reliability of a Real-Time Redundant Computational Process in Cluster Architecture Systems Automatic Control and Computer Sciences, Vol. 49, No. 1, 2015, pp. 46-56.
7. V.A.Bogatyrev Exchange of Duplicated Computing Complexes in Fault tolerant Systems Automatic Control and Computer Sciences Vol. 46, No. 5, 2011, pp. 268-276.
8. V.A.Bogatyrev Fault tolerance of clusters configurations with direct connection of storage devices Automatic Control and Computer Sciences V. 45, No. 6, 2011, pp. 330-337.
9. 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 - 2016, Vol. 601, pp. 199-207.
10. V.A.Bogatyrev An interval signal method of dynamic interrupt handling with load balancing Automatic Control and Computer Sciences, vol. 34 No. 6, 2000, pp. 51-57.
11. V.A.Bogatyrev Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer Automatic Control and Computer Sciences, vol. 33, No. 1, 1999, pp. 57-63.
12. V.A.Bogatyrev On interconnection control in redundancy of local network buses with limited availability Engineering Simulation, Vol. 16, No 4, 1999, pp. 463-469.
13. Aliev T. I. The Synthesis of Service Discipline in Systems with Limits // Distributed Computer and Communication Networks 18th International Conference, DCCN 2015. - Moscow, Russia, October 19-22, 2015. Revised Selected Papers, pp/151-156.
14. V.S.Kolomoitcev, V.A.Bogatyrev Selecting multilevel structure secure access to resources external network Distributed computer and communication networks: control, computation, communications (DCCN-2015) Conf., 2015, pp. 525-532.

УДК 517.937, 517.928.2, 519.216.2

Анализ решений сингулярно возмущенного стохастического дифференциального уравнения бесконечного порядка

Е. О. Коршок, С. А. Васильев

*Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

Аннотация. В данной работе предлагается алгоритм построения асимптотических решений сингулярно возмущенного стохастических дифференциального уравнения бесконечного порядка.

Ключевые слова: стохастические дифференциальные уравнения, дифференциальные уравнения бесконечного порядка, сингулярно возмущенные дифференциальные уравнения.

1. Введение

Применение при моделировании динамики сложных систем стохастических дифференциальных уравнений (СДУ) представляет подход, позволяющий решить многие задачи: технические, экономические и др. В большинстве существующих подходов применяются методы, позволяющие получить решение-траекторию путем сведения задач большой размерности к серии задач меньшей размерности [1], [2], [3], [4], [5], [6], [7], [8], [9], также исследовались стохастические дифференциальные уравнения бесконечного порядка [14]. Дифференциальные уравнения бесконечного порядка рассматривались в работах А.Н.Тихонова [15], К.П. Персидского [13], О.А. Жаутыкова [18], [19], Ю.Ф. Коробейника [10] и другие. Также большой интерес представляют работы в области сингулярно возмущенных дифференциальных уравнений: А.Н.Тихонова [16], А.Б. Васильевой [17], С.А. Ломова [12] и другие.

В данной работе предлагается алгоритм построения асимптотических решений сингулярно возмущенного стохастических дифференциального уравнения бесконечного порядка и исследуется вопрос о существовании и единственности его решения.

2. Постановка задачи

2.1. Сингулярно возмущенное стохастическое дифференциальное уравнение бесконечного порядка

Рассмотрим сингулярно возмущенное стохастическое дифференциальное уравнение

$$\mu^{-1} \left(\exp \left(\mu \frac{dX_t}{dt} \right) - 1 \right) = b(t, X_t) + \sigma(t, X_t)W_t, \quad (1)$$

где X_t – состояние системы в момент времени t , функции $b(t, X_t) \in \mathbf{R}$, $\sigma(t, X_t) \in \mathbf{R}$, а W_t – мерный белый шум, $\mu \in (0, 1]$ – малый параметр.

Уравнение (1) можно представить в виде стохастического дифференциального уравнения бесконечного порядка:

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{k!} \frac{d^k X_t}{dt^k} &= b(t, X_t) + \sigma(t, X_t)W_t = \\ &= \frac{dX_t}{dt} + \frac{\mu}{2!} \frac{d^2 X_t}{dt^2} + \frac{\mu^2}{3!} \frac{d^3 X_t}{dt^3} + \dots = b(t, X_t) + \sigma(t, X_t)W_t, \end{aligned}$$

но, с другой стороны, уравнение (1) можно записать таким образом, что оно будет конечно-разностным:

$$\frac{X_t(t - \mu)}{\mu} = b(t, X_t) + \sigma(t, X_t)W_t, \quad (2)$$

где $\mu \ll 1$ – сдвиг во времени.

Если в этом уравнении формально устремить величину $\mu \rightarrow 0$, то уравнение (2) переходит в уравнение

$$\frac{X_t}{dt} = b(t, X_t) + \sigma(t, X_t)W_t,$$

которое будем называть вырожденным.

Для удобства перепишем уравнение (1) в виде:

$$\tilde{L}_{\infty}^{\mu} X_t(t, \mu) = 0, \tilde{L}_{\infty}^{\mu} = L + \mu L_{\infty}^{\mu}; \quad (3)$$

где оператор \tilde{L}_{∞}^{μ}

$$L = \frac{d}{dt} - b - \sigma W_t; L_{\infty}^{\mu} = \sum_{k=2}^{\infty} \mu^{k-2} \frac{\mu^{k-1}}{k!} D^k X_t, D^k = \frac{d^k}{dt^k}.$$

Для уравнения (3) сформулируем задачу Коши:

$$\begin{cases} \tilde{L}_{\infty}^{\mu} X_t(t, \mu) = 0, \\ D^n X_0(0) = \bar{X}_n, n = 0, 1, \dots, \end{cases} \quad (4)$$

где $\bar{X}_n \subset l_1$ – числовая последовательность, определяющая начальное состояние системы.

Задача Коши (4) является задачей с малым параметром ($\mu \ll 1$) при старших производных и поэтому ее можно отнести к классу сингулярно возмущенных задач, так как при формальном устремлении $\mu \rightarrow 0$ (4) порядок дифференциального уравнения понизится, задача станет вырожденной и переопределенной; в связи с этим встанет о выборе начальных условий для вырожденной задачи.

Для уравнения (4) сформулируем вырожденную задачу Коши:

$$\begin{cases} L \tilde{X}_t = 0, \\ \tilde{X}_0(0) = \bar{X}_0, \end{cases} \quad (5)$$

где $\bar{X}_0 \in R$ – действительная величина, определяющая начальное состояние системы. Таким образом, возникает вопрос о построении асимптотического решения задачи (4) и о выборе начальных условий для вырожденной задачи (5).

2.2. Усечение сингулярно возмущенного стохастического дифференциального уравнения

Если в уравнении (3) ограничится конечным порядком $m > 1$, тогда его можно записать таким образом:

$$\begin{aligned} \tilde{L}_m^\mu X_t^m(t, \mu) = 0, \quad \tilde{L}_m^\mu = L + \mu L_m^\mu; \quad (6) \\ L = \frac{d}{dt} - b^m - \sigma^m W_t^m; \quad L_m^\mu = \sum_{k=2}^m \mu^{k-2} \frac{\mu^{k-1}}{k!} \frac{d^k X_t}{dt^k}. \end{aligned}$$

Для уравнения этого уравнения сформулируем задачу Коши:

$$\begin{cases} \tilde{L}_m^\mu X_t^m(t, \mu) = 0, \\ D^n X_0^m = \bar{X}_n^m, \quad n = 0, 1, \dots, m \end{cases} \quad (7)$$

где $\bar{X}_n^m \in \mathbf{R}$ – m действительных чисел, определяющих начальное состояние системы.

Для уравнения (6) сформулируем вырожденную задачу Коши:

$$\begin{cases} L \tilde{X}_t = 0, \\ \tilde{X}_0 = \bar{X}_0, \end{cases}$$

где $\bar{X}_0 \in R$ – действительная величина, определяющая начальное состояние системы. Будем считать, что данная задача совпадает с задачей (5)

Задача Коши (7) является задачей с малым параметром ($\mu \ll 1$) при старших производных и поэтому ее можно отнести к классу сингулярно возмущенных задач, так как при формальном устремлении $\mu \rightarrow 0$ (7) порядок дифференциального уравнения понизится, задача станет вырожденной и переопределенной; в связи с этим встанет о выборе начальных условий для вырожденной задачи.

Таким образом, возникает вопрос о построении асимптотического решения задачи (7) и о выборе начальных условий для вырожденной задачи (5).

3. Формализм построения асимптотического решения задачи Коши для СДУ

3.1. Разложение по малому параметру

Будем искать формальное решение X_t задачи Коши (4) в виде такого асимптотического ряда:

$$\Theta X_t = \bar{X}_t + \Pi X_t = \sum_{k=0}^{\infty} \mu^k (\bar{X}_{tk} + \Pi X_{tk}), \quad (8)$$

что его частичная сумма

$$\Theta_j X_t = \sum_{k=0}^j \mu^k (\bar{X}_{tk} + \Pi X_{tk})$$

будет удовлетворять неравенствам для решения задачи (4)

$$\max_{t \in [\delta_A, t_0 - \delta_A]} |X_t - \Theta_j X_t| < M \mu^{j+1}$$

а также аналогичным неравенствам для краевых условий данных задач, где M , и $\delta \ll 1$, — положительные постоянные, независимые от t и μ . Тогда для X_t асимптотическое решение будет иметь вид:

$$X_t = \sum_{k=0}^j \mu^k (\bar{X}_{tk} + \Pi X_{tk}) + \bar{Z}_j(t),$$

где $\bar{Z}_j(t) = \mu^{j+1} z_j(t)$ — погрешность асимптотического приближения решения X_t частичной суммой $\Theta_j X_t$

$$\bar{Z}_j(t) = X_t - \Theta_j X_t.$$

Здесь $\bar{X}_t(t, \mu)$ — регулярная часть разложения, а $\Pi X_t(\tau, \mu)$ — пограничная функция, описывающий поведение решения на $t \in [0, t_0]$, $t_0 >$

0. Для пограничной функции $\Pi X_t(\tau, \mu)$ здесь введена новая независимая ("растянутая") переменная $\tau = t/\mu$.

Кроме того, будем предполагать возможность разложения функции b , σ в виде сходящихся рядов в окрестности точки $t = 0$

$$\begin{aligned} b(t) &= \sum_{k=0}^{\infty} \mu^k b_k, \quad \sigma(t) = \sum_{k=0}^{\infty} \mu^k \sigma_k, \\ b(\tau) &= \sum_{k=0}^{\infty} \mu^k \Pi b_k, \quad \sigma(\tau) = \sum_{k=0}^{\infty} \mu^k \Pi \sigma_k, \end{aligned} \quad (9)$$

3.2. Члены асимптотики

Подставим разложения (8)-(9) в уравнение и краевые условия задачи Коши (4) и приравняем члены, стоящие при одинаковых степенях μ , таким образом, чтобы получить краевые задачи для определения членов разложения (8) соответствующей задачи.

При этом на пограничную функцию ΠX_{tk} мы накладываем такие дополнительные условия, которые обеспечивают стремление этой функций к нулю вне пограничного слоя, т. е. $\Pi X_{tk}^{\infty} \rightarrow 0$, $k = 0, 1, 2, \dots$ при $\mu \rightarrow 0$ и фиксированном t .

В нулевом приближении мы получим систему такого вида:

$$\begin{cases} L\bar{X}_{t0} = 0, \\ \bar{X}_{t0}(0) = \bar{X}_0, \end{cases}$$

которая совпадает с задачей (4). В первом приближении система выглядит так:

$$\begin{cases} L\bar{X}_{t1} = -\frac{1}{2!} D^2 \bar{X}_{t0}, \\ L_{\infty}^{\Pi} \Pi X_{t1} = 0, \\ D^n (\bar{X}_{t1}(0) + \Pi X_{t1}(0)) = \bar{X}_n, \\ \Pi X_{t1}(\tau) \rightarrow 0, \\ n = 0, 1, 2, \dots \end{cases}$$

$$L_{\infty}^{\Pi} = \sum_{k=1}^{\infty} \frac{1}{p!} \frac{d^p}{d\tau^p}.$$

В случае $k > 1$ для задач (4) мы получим системы уравнений и дополнительные условия для нахождения такого вида:

$$\begin{cases} L\bar{X}_{tk} = -h_k, \\ L_{\infty}^{\Pi} \Pi X_{tk} = \Pi b_{k-1} + \Pi \sigma_{k-1} W_t, \\ D^n(\bar{X}_{tk}(0) + \Pi X_{tk}(0)) = \bar{X}_n, \\ \Pi X_{t1}(\tau) \rightarrow 0, \\ n = 0, 1, 2, \dots, \\ h_k = \sum_{p=1}^k \frac{1}{p!} D^p \bar{X}_{tk-p}. \end{cases}$$

Таким образом, описанный алгоритм позволяет найти асимптотическое решение задач (4) для любого порядка j . Аналогичные выкладки можно провести для задачи Коши (7). В нулевом приближении мы получим систему такого вида:

$$\begin{cases} L\bar{X}_{t0}^m = 0, \\ \bar{X}_{t0}^m(0) = \bar{X}_0, \end{cases}$$

которая совпадает с задачей (4). В первом приближении система выглядит так:

$$\begin{cases} L\bar{X}_{t1}^m = -\frac{1}{2!} D^2 \bar{X}_{t0}^m, \\ L_m^{\Pi} \Pi X_{t1}^m = 0, \\ D^n(\bar{X}_{t1}^m(0) + \Pi X_{t1}^m(0)) = \bar{X}_n, \\ \Pi X_{t1}^m(\tau) \rightarrow 0, \\ n = 0, 1, 2, \dots \\ L_m^{\Pi} = \sum_{k=1}^m \frac{1}{k!} \frac{d^k}{d\tau^k}. \end{cases}$$

В случае $k > 1$ для задач (7) мы получим системы уравнений и дополнительные условия для нахождения такого вида:

$$\begin{cases} L\bar{X}_{tk}^m = -h_k^m, \\ L_m^{\Pi} \Pi X_{tk}^m = \Pi b_{k-1} + \Pi \sigma_{k-1} W_t, \\ D^n(\bar{X}_{tk}^m(0) + \Pi X_{tk}^m(0)) = \bar{X}_n, \\ \Pi X_{t1}^m(\tau) \rightarrow 0, \\ n = 0, 1, 2, \dots, \end{cases}$$

где при $k \leq m$

$$h_k^m = \sum_{p=1}^k \frac{1}{p!} D^p \bar{X}_{t k-p}^m,$$

при $k > m$

$$h_k^m = \sum_{p=1}^m \frac{1}{p!} D^p \bar{X}_{t k-p}^m.$$

4. Заключение

В данной работе представлен алгоритм построения асимптотических решений сингулярно возмущенного стохастических дифференциального уравнения бесконечного порядка. На основе этого алгоритма имеется возможность нахождения асимптотического приближения решения задачи Коши для сингулярно возмущенного стохастических дифференциального уравнения как бесконечного порядка, так и конечного порядка m , что позволяет в дальнейшем применить численные алгоритмы для приближенного поиска решения таких уравнений.

Благодарности

Данное исследование выполнено на основании договора No 02.a03.21.0008 от 24.11.2015 между Министерством образования и науки РФ и РУДН.

Литература

1. *Berglund N., Gentz B.* Geometric singular perturbation theory for stochastic differential equations // Journal of Differential Equations. — 2003. — Vol. 191, No 1. — С.1–54.
2. *Carroll C., Tokuoka K., Wu W.* The Method of Moderation for Solving Dynamic Stochastic Optimization Problems. — Paper provided by Society for Economic Dynamics in its series 2012 Meeting Papers with number 1102..
3. *Marti K.* Stochastic optimization methods. — Springer, Berlin Heidelberg, 2005. — ISBN: 978-3-662-46214-0.
4. *Кабанов Ю.М., Пергаменичиков С.М.* Сингулярные возмущения стохастических дифференциальных уравнений // Матем. сб. — 1990. — Том 181, No 9. — С.1170–1182.
5. *Кабанов Ю.М., Пергаменичиков С.М.* О сингулярно возмущенных стохастических дифференциальных уравнениях и уравнениях в частных производных // ДАН СССР. — 1990. — Том 311, No 5. — С.1039 – 1042.
6. *Kabanov Yu.M., Pergamenshchikov S.M.* Optimal control of singularly perturbed linear stochastic systems // Stochastics and Stoch. Rep. — 1991. — Vol.36 — С.109 – 135.

7. *Kabanov Yu.M., Pergamenschikov S.M., Stoyanov J.M.* Asymptotic expansions for singularly perturbed stochastic differential equations // *Stochastics and Stoch. Rep. – New Trend in Probability and Statistics, Proc. of the Bakuriani Coll. in Honour Yu.?V. Prokhorov. V. 1, eds. V.?V. Sazonov, T.?L. Shervashidze, Mokslas, Vilnius; VSP, Utrecht, 1991, pp. 413 – 435.*
8. *Пергаменщиков С.М.* Асимптотические разложения для моделей с быстрыми и медленными переменными, описываемые сингулярно возмущенными системами стохастических дифференциальных уравнений // *УМН – 1994. – Том 49, No 4. – С.3 – 46.*
9. *Stein, Jerome L.* *Stochastic Optimal Control, International Finance, and Debt Crises.* – Oxford University Press, 2006. – ISBN: 978-0-199-28057-5.
10. *Korobeinik Ju.* Differential equations of infinite order and infinite systems of differential equations. *Izv. Akad. Nauk SSSR Ser. Mat. Vol. 34, pp. 881 – 922 (1970)*
11. *Krasnoselsky M.A., Zabreyko P.P.* *Geometrical methods of nonlinear analysis.* Springer-Verlag, Berlin (1984)
12. *Lomov S. A.* The construction of asymptotic solutions of certain problems with parameters. *Izv. Akad. Nauk SSSR Ser. Mat. Vol. 32, pp. 884 – 913 (1968)*
13. *Persidsky K.P.* *Izv. AN KazSSR, Ser. Mat. Mach., Issue 2, pp. 3 – 34. (1948)*
14. *Skorokhod A.* On infinite systems of stochastic differential equations // *Methods Funct. Anal. Topology. Vol. 5, No. 4, 54 – 61 (1999)*
15. *Tihonov A. N.* *Uber unendliche Systeme von Differentialgleichungen. Rec. Math. Vol. 41, Issue 4, pp. 551 – 555 (1934).*
16. *Tihonov A. N.* Systems of differential equations containing small parameters in the derivatives. *Mat. Sbornik N. S. Vol. 31, Issue 73, pp. 575 – 586 (1952)*
17. *Vasil'eva A. B.* Asymptotic behaviour of solutions of certain problems for ordinary non-linear differential equations with a small parameter multiplying the highest derivatives. *Uspehi Mat. Nauk. Vol. 18, Issue 111, no. 3, 15 – 86 (1963)*
18. *Zhautykov O. A.* On a countable system of differential equations with variable parameters. *Mat. Sb. (N.S.). Vol. 49, Issue 91, pp. 317 – 330 (1959)*
19. *Zhautykov O. A.* Extension of the Hamilton-Jacobi theorems to an infinite canonical system of equations. *Mat. Sb. (N.S.). Vol. 53, Issue 95, pp. 313 – 328 (1961)*

UDC 517.937, 517.928.2, 519.216.2

Solutions analysis of infinite order singular perturbed stochastic differential equation

E. O. Korshok, S. A. Vasilyev

*Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

In this paper, we propose an algorithm for asymptotic solutions of a singularly perturbed stochastic differential equations of infinite order. Solutions of singularly perturbed stochastic differential equations of infinite order was built.

Keywords: stochastic differential equations, infinite order differential equations; singular perturbed differential equations.

UDC 519.8, 004.7

On Convergence of Active Control Strategies for Subscription Notification Delivery in Smart Spaces

D. G. Korzun, A. S. Vdovenko, O. I. Bogoiavlenskaia

*Department of Computer Science
Petrozavodsk State University*

Lenin Ave. 33, Petrozavodsk, Republic of Karelia, 185910, Russia

Abstract. A smart space provides a shared view on information, which is cooperatively produced, processed, and consumed by participants themselves in a computing environment. One of the most advanced networked operations on this information is the subscription operation. It supports the information-driven programming style: a notification is delivered to all interested participants when an appropriate information fact is formed in the smart space. In this work, we continue our study of the notification delivery when notifications assigned to a mobile participant are undelivered though several information updates have been made by others. The notification delivery performance is improved by active control: the participant proactively tracks the information formation in accordance to individually and rationally defined time points. We analytically and experimentally evaluate the convergence of such a strategy for different distributions of the notification loss.

1. Introduction

The emerging technologies of Internet of Things (IoT) lead now to the new type of ubiquitous computing environments (IoT environments) where the role of distributed processing of the information from multiple available sources by multiple participants is essential [2,5]. Participants are represented by various digital devices, among which a large and growing family is personal mobile devices that represent end-users. A promising paradigm for programming such an IoT environment is smart spaces [1,7]. A smart space is deployed in a given computing environment and provides a shared view on information, which is cooperatively produced, processed, and consumed by participants themselves.

The publish/subscribe (pub/sub) model is widely used for multi-party interaction in distributed systems [3]. In smart spaces, the subscription operation is one of the most advanced networked operations on the shared information [9,11]. The operation supports the information-driven programming style: a notification is delivered to all interested participants when an appropriate information fact is formed in the smart space.

In many existing solutions for smart spaces, the major role in information sharing is played by a semantic information broker (SIB) [4,7]. In particular, SIB is responsible for detection of information changes and for subsequent delivery of notifications to those clients that subscribed to the

information. Both change detection and notification delivery are subject to losses in network environments [12].

We continue our study of an active control of notification delivery for subscription operation for the case of notification losses. The problem was initially stated in [12]: the client follows an adaptive strategy controlling the check interval based on the number of notifications lost in the latest window. This adaptive strategy was early experimented in [6] and we showed that the strategy is a generalization of the TCP algorithm of additive-increase/multiplicative-decrease (AIMD). In this paper, we analyze certain convergence properties of the adaptive strategy compared with the multiplicative-decrease strategy and for different distributions of notification losses. Our study is focused on the typical case for smart spaces with many mobile end-users who concurrently make changes to the shared information as well as interested in detection of such changes [8].

The rest of the paper is organized as follows. Section 2 formulates the notification loss problem for the subscription operation in smart spaces. Section 3 introduces two control strategies for convergence analysis: the adaptive strategy, which is a generalization of the AIMD algorithm of TCP, and the multiplicative-decrease strategy. Section 4 presents results of our convergence analysis, both theoretical and simulation estimates, for the two notification loss distributions (uniform and Poisson losses). Section 5 concludes the paper.

2. Notification Delivery Problem

A smart space forms a sparse-connected multi-agent system deployed in a given IoT environment [7]. Such an environment consists of various digital devices, which act as IoT smart objects [5]. Software agents run on the devices and interact over the shared information content. Some agents represent end-users using personal mobile devices (e.g., smartphone) This type of interaction involves, in parallel and asynchronously, a lot of informational sources and destinations. Information sharing makes the interaction indirect, based on a semantic information broker (SIB) [8, 11]. The latter implements a shared information storage, serving requests from agents on read/write operations.

The subscription operation specifies a persistent query from an agent (a subscription client) to the SIB (a subscription server) for a particular part of the shared content [9]. Whenever the specified part is changed, the agent should receive the subscription notification. Changes are due to parallel activity of other agents, which act as publishers in this interaction (note that an agent may combine the roles of publisher and subscriber). SIB monitors subscriptions of all clients and maps all content changes to the specified interests. Therefore, changes are controlled on the SIB side, and corresponding notifications are sent to the clients. SIB acts as a passive receiver, and we call such subscription notifications passive [12].

We employ Smart-M3 as a reference software platform for creating smart spaces [4, 8]. For each subscription, the SIB maintains a network connection (e.g., a TCP connection) established by the client's request [9]. Knowing the set of all subscriptions, the SIB regularly checks that they are alive, removing the subscription if its network connection is lost.

Smart-M3 follows the best effort style in subscription notification delivery. A notification *should* be sent to a client if a related change in the content has happened. Some notifications can be unsent by SIB due to its overload or internal operability faults. SIB does not check delivery for already sent notifications, and a new notification can be sent although the underlying network connection is broken on the client side.

The above properties do not ensure the dependable notification delivery even if reliable network protocols are used, such as TCP. A possible solution for a client is augmenting the passive notification delivery with an active control strategy to reduce the number of undelivered notifications.

Consider the following model to formalize the key properties of the subscription notification loss problem in smart spaces. Let $i = 1, 2, \dots$ be the event-based time evolution on the client side, where i is the index of notification events. An event i is either a passive notification (i.e., received from SIB) or an explicit check of the notification delivery (made by the client within its active control). Denote by t_i and k_i the time elapsed and the number of losses occurred between i and $i + 1$, respectively. Assume that some initial value t_0 is always defined. The values for k_i are non-negative integers.

3. Active Control Strategy Models

Following our previous work [6] we consider the following two active control strategies that implements "adaptation to losses" for the client. This kind of adaptation means that the client reduces its check interval t_i when losses are observed, and increases t_i in the case of no losses.

Adaptive strategy Let the client has observed no losses during t_{i-1} , i.e., $k_{i-1} = 0$. The observation indicates the system state. The client increases additively the check interval, i.e., $t_i = t_{i-1} + \delta$ for a fixed parameter $\delta > 0$. On the contrary, if the client has observed losses, i.e., $k_{i-1} > 0$, it reduces t_i to decrease the number of losses in the nearest future. The reduction applies the multiplicative average

$$t_i = \alpha t_{i-1} + (1 - \alpha) \frac{t_{i-1}}{k_{i-1} + 1}$$

for a fixed parameter $0 \leq \alpha < 1$.

We yield the recurrent system by which t_i is reduced (multiplicative decrease) in case of losses and incremented (additive increase) otherwise:

$$t_i = \begin{cases} t_{i-1} + \delta & \text{if } k_{i-1} = 0, \\ \frac{1 + \alpha k_{i-1}}{k_{i-1} + 1} t_{i-1} & \text{if } k_{i-1} > 0. \end{cases} \tag{1}$$

Note that (1) is valid only for active control. When a passive notification i is delivered, then the value of t_i is not set by the client.

Multiplicative–decrease strategy It is a semi-adaptive approach, in which the check interval is halved in case of losses and set to the initial (reference) value t_0 otherwise:

$$t_i = \begin{cases} t_{i-1}/2 & \text{if } k_{i-1} > 0, \\ t_0 & \text{if } k_{i-1} = 0. \end{cases} \tag{2}$$

For the multiplicative–decrease strategy its evolution (2) can be described by a discrete-time Markov chain. Since the check interval accepts one of the values $t_0/2^i$ for $i = 0, 1, \dots$, we consider it a random variable. Assume that the sequence $\{i_n\}_{n \geq 0}$ of the indices of 2 forms a Markov chain which is a random walk with probabilities $p_{i,i+1} = q_i$ and $p_{i0} = p_i$. Denote $q_i = 1 - p_i$ for $i = 0, 1, \dots$. According to [10] the chain has a steady state distribution $\pi_i = \lim_{n \rightarrow \infty} p_{ij}$, if the following condition holds

$$\prod_{k=0}^{\infty} q_k = \lim_{n \rightarrow \infty} q_0 q_1 \cdots q_n = 0.$$

4. Convergence Evaluation

In the subsequent analysis we consider the following distributions to model the notification losses, in accordance with [6].

1. Let the time elapsed between consecutive losses follow a uniform distribution $\mathcal{U}\{0, \xi t_0\}$. Hence, the average number of losses in any check interval is proportional to its length t_i

2. Let k_i follow a Poisson process of parameter λ . Hence, the number of losses during t_i has the probability mass function

$$\mathbb{P}(k_i = k) = \frac{(\lambda t_i)^k}{k!} e^{-\lambda t_i}$$

Convergence of the multiplicative–decrease strategy to a steady state distribution For practical use, the convergence to the steady state distribution needs to be exponential in dependence on the number of steps n . For analytical estimation, we apply the following criterion described in [10]. The ergodicity coefficient is defined as

$$k(n_0) = 1 - \frac{1}{2} \sup_{i,i+1} \sum_{m=0}^{\infty} |p_{im}(n_0) - p_{jm}(n_0)|,$$

where $p_{ij}(n_0)$ is the transition probability from the state i to the state j in n_0 time steps of the Markov chain.

Let us assume $q_i \leq 1 - \Delta < 1$, i.e., the transition probability from any state $i \neq 0$ into the zero state satisfies the inequality $p_i \geq \Delta$. Therefore one can obtain the following estimation for the ergodicity coefficient $k(1)$:

$$k(1) \geq \inf_i p_i \geq \Delta.$$

Thus the convergence rate of the Markov chain under consideration to the steady state distribution $\{\pi_j\}_{j=0}^{\infty}$ could be estimated as

$$\sup_j |\pi_j(n) - \pi_j| \leq (1 - k(1))^{n-1},$$

where $\pi_j(n)$ is the probability that the chain is in the state j at the step n .

For the Poisson notification losses $p_0 = e^{-\lambda t_0}$ and $p_n = p_0^{2^{-n}}$; it yields

$$\inf_i p_i = \inf_i (e^{-\lambda t_0})^{2^{-i}} = \inf_i e^{-\frac{\lambda t_0}{2^i}} \geq e^{-\lambda t_0}.$$

Consequently,

$$\sup_j |\pi_j(n) - \pi_j| \leq (1 - e^{-\lambda t_0})^{n-1}.$$

As a result, the Markov chain convergence rate is slower for higher initial values t_0 . In our simulation settings (see below), the difference becomes smaller 0.05 on step $n = 8$. Note that too small values of t_0 lead to performance degradation due to frequent checks.

For the uniform distribution one obtains $q_n = q_0 2^{-n}$ and correspondingly $p_n = 1 - q_0 2^{-n}$. Consequently, $\inf_i (1 - q_0 2^{-i}) \geq 1 - q_0$, and

$$\sup_j |\pi_j(n) - \pi_j| \leq (1 - 1 - q_0)^{n-1} = q_0^{n-1}.$$

As a result, the Markov chain convergence speed is slower for higher values of the probability q_0 . Assuming $q_0 = 0.3$ the difference becomes smaller 0.05 on step $n = 4$.

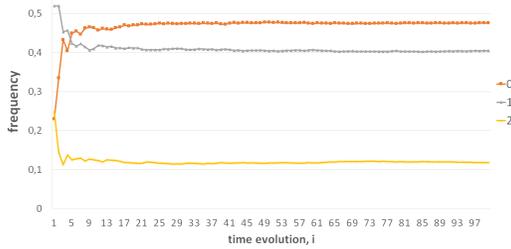


Figure 1. Experimental convergence of l_k for the multiplicative–decrease strategy and uniform notification losses ($k = 0, 1, 2$).

Table 1
Performance comparison of the experimented strategies

Strategy	Metric	Distribution	
		Poisson	Uniform
Multiplicative-decrease strategy	k_{avg}	0.729	0.646
	t_{avg}	14.382	13.959
	$k_{\text{avg}}/t_{\text{avg}}$	0.051	0.046
Adaptive strategy	k_{avg}	1,138	0,996
	t_{avg}	22,989	20,838
	$k_{\text{avg}}/t_{\text{avg}}$	0,049	0,048

Simulation comparison of the convergence for the adaptive and multiplicative–decrease strategies Denote $l_k = \mathbb{P}[k_i = k]$ the probability that the number of losses is k on some interval t_i . We performed simulation experiments for the two active control strategies and for the two notification loss distributions. We estimate l_k as a frequency converging to theoretical values for the probabilities $\mathbb{P}[k_i = k]$. The results are visualized in Figure 1 for uniform distribution of losses.

For uniform losses we set $\xi = 0.1$. For Poisson losses we set $\lambda = 0.05$. In both active strategies, we take $t_0 = 20$ s. According to the selected parameters of the loss distributions, it means that every 20 seconds one notification is lost on average. For strategy (1) we use $\alpha = 0.3$ and $\delta = t_0$.

The multiplicative–decrease strategy for uniform losses (Figure 1) has low values of losses (i.e., $k_i \leq 2$). In contrast, for Poisson losses the number of simultaneous losses per interval can be high (i.e., $k_i = 6$ in some cases). The adaptive strategy has higher number of losses since it can set the check interval t_i more than t_0 .

The experiments show that the probabilities l_k are decreasing in dependence on $k = 0, 1, \dots$. As a result, many simultaneous losses happen rare. The probability of no loss is highest and lies in $0.4 \leq l_0 \leq 0.5$, which means that the strategies keep about half of check intervals in the “no loss state”. Clearly, l_0 can be made higher by setting lower t_0 . The convergence of l_k is fast for both strategies and all k .

Table 1 shows quantitative comparison based on the efficiency metrics from [6]. Let $k_{\text{avg}} \approx \mathbb{E}[K]$ be the number of losses per interval on average and $t_{\text{avg}} \approx \mathbb{E}[T]$ be the average length of check interval:

$$k_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n k_i, \quad t_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n t_i.$$

The ratio $k_{\text{avg}}/t_{\text{avg}}$ shows the strategy performance: the lower the metric value the higher the performance.

The experimental evaluation shows that the strategies have similar performance in the terms of metric $k_{\text{avg}}/t_{\text{avg}}$. This result is mainly due to the static distribution of losses and a priori selection of t_0 . Nevertheless, the adaptive strategy provides faster convergence to the steady state.

5. Conclusion

This paper continues our study of active control strategies for a mobile client subscribed to certain information in a smart space. Such strategies achieve better performance in the terms of ratio “the number of losses vs. the check interval”. Our analytical and experimental evaluation indicates that the convergence speed to the steady state is reasonable for smart spaces applications in IoT environments, where many mobile users can concurrently make changes to the shared information as well as they are interested in dependable detection of such changes.

Acknowledgments

This research is financially supported by the Ministry of Education and Science of the Russian Federation within project # 1481 of the basic part of state research assignment for 2014–2016. The reported study was partially funded by RFBR according to research project # 14-07-00252.

References

1. Augusto, J., Callaghan, V., Cook, D., Kameas, A., Satoh, I.: Intelligent environments: a manifesto. *Human-centric Computing and Information Sciences* 3(1), 1–18 (2013)

2. Cristea, V., Dobre, C., Pop, F.: Internet of Things and Inter-cooperative Computational Technologies for Collective Intelligence, chap. Context-Aware Environments for the Internet of Things, pp. 25–49. Springer Berlin Heidelberg (2013)
3. Eugster, P.T., Felber, P.A., Guerraoui, R., Kermarrec, A.M.: The many faces of publish/subscribe. *ACM Comput. Surv.* 35, 114–131 (June 2003)
4. Honkola, J., Laine, H., Brown, R., Tyrkkö, O.: Smart-M3 information sharing platform. In: *Proc. IEEE Symp. Computers and Communications (ISCC'10)*. pp. 1041–1046. IEEE Computer Society (Jun 2010)
5. Kortuem, G., Kawsar, F., Sundramoorthy, V., Fitton, D.: Smart objects as building blocks for the Internet of Things. *IEEE Internet Computing* 14(1), 44–51 (Jan 2010)
6. Korzun, D., Pagano, M., Vdovenko, A.: Control strategies of subscription notification delivery in smart spaces. In: Vishnevsky, V., Kozyrev, D. (eds.) *Distributed computer and communication networks. CCIS 601*, pp. 40–51. Springer International Publishing (2016)
7. Korzun, D., Balandin, S., Gurtov, A.: Deployment of Smart Spaces in Internet of Things: Overview of the design challenges. In: Balandin, S., Andreev, S., Koucheryavy, Y. (eds.) *Proc. 13th Int'l Conf. Next Generation Wired/Wireless Networking and 6th Conf. on Internet of Things and Smart Spaces, LNCS 8121*. pp. 48–59. Springer (Aug 2013)
8. Korzun, D., Kashevnik, A., Balandin, S., Smirnov, A.: The Smart-M3 platform: Experience of smart space application development for Internet of Things. In: Balandin, S., Andreev, S., Koucheryavy, Y. (eds.) *Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Proc. 15th Int'l Conf. Next Generation Wired/Wireless Networking and 8th Conf. on Internet of Things and Smart Spaces, LNCS 9247*. pp. 56–67. Springer (Aug 2015)
9. Lomov, A.A., Korzun, D.G.: Subscription operation in Smart-M3. In: Balandin, S., Ovchinnikov, A. (eds.) *Proc. 10th Conf. of Open Innovations Association FRUCT and 2nd Finnish–Russian Mobile Linux Summit*. pp. 83–94. SUAI (Nov 2011)
10. Rozanov, Y.: *Sluchaynyi protsessyi. Kratkiy kurs.* (in Russian). Nauka, 2nd edn. (1979)
11. Smirnov, A., Kashnevik, A., Shilov, N., Oliver, I., Balandin, S., Boldyrev, S.: Anonymous agent coordination in smart spaces: State-of-the-art. In: *Smart Spaces and Next Generation Wired/Wireless Networking. Proc. 9th Int'l Conf. NEW2AN'09 and 2nd Conf. on Smart Spaces ruSMART 2009. LNCS 5764*. pp. 42–51. Springer-Verlag, Berlin, Heidelberg (2009)
12. Vdovenko, A., Korzun, D.: Active control by a mobile client of subscription notifications in smart space. In: *Proc. 16th Conf. Open Innovations Framework Program FRUCT*. pp. 123–128 (Oct 2014)

UDC 621.395

Applications of augmented reality traffic and quality requirements study and modeling

A. Koucheryavy, M. Makolkina, A. Paramonov

*The Bonch-Bruевич Saint-Petersburg State University of Telecommunications,
Saint-Petersburg, Russia*

Abstract. The further development of communication networks appears today on the basis of the concept of the Internet of Things. At the same time gaining popularity technology "augmented reality" that allows you to manage different objects and processes in networks. Sharing the "augmented reality" and the concept of the Internet of Things technology, requires the development of the new service model and traffic pattern and establishment of a new approach of the Quality of Experience estimation.

Keywords: augmented reality, Internet of Things, Quality of Experience, service model, traffic pattern, D2D technology.

1. Introduction

Today in the market of telecommunications there was a new type of service - it is services augmented reality (AR, Augmented Reality) [1, 2]. Primal problem of technology augmented reality not to create the new world, and to improve existing for the account strengthenings of feelings of the user such as hearing, vision, sense of smell, knowledge, etc. The augmented and virtual realities in essence different concepts. The virtual reality replaces the actual world synthetic, artificially created environment, being in which, the person does not see what occurs around. At that time as AR supplements a real, but does not replace it. It allows the user to see the actual world with the virtual objects combined or imposed atop actual world [3, 4]. Each person, not very well who is he, the doctor or the climber, in search of the nearest subway station in the big city will feel all advantages of new technology which allows to impose computer graphics on environmental space under review the person, and by that gives the chance to glance in buildings and to see through walls. So, walking down the street augmented reality wearing glasses which look as the routine couple of points, is available to you graphics with various information in dependence from your turn of the head and orientation of a look, audio, video, data on objects coincide with what you see at present. So, the main idea of augmented reality consists in imposing of graphics, audio, other sensory data and feelings over the existing world around in real time scale [5]. Thus, it is possible to mark out three main properties AR technologies:

- 1) A combination of real and virtual objects in uniform space;
- 2) Interactivity in real time;
- 3) 3D objects.

In the modern applications of augmented reality it is possible to allocate on to smaller measure six classes: medicine; assembly, maintenance and repair the difficult technique; addition of information of private and common character to existing to objects; control of robots, unmanned aerial vehicles, etc.; games and entertainments; military.

2. Applications of augmented reality

The range of application of technology of augmented reality in medicine is extensive and bears a number of benefits for doctors in the composite cases demanding fast reaction and acceptance decisions. As it is known a lot of time leaves on diagnostics, poll and survey the patient, here the doctor possesses as if "x-ray vision", and looking on particular parts of a body of the patient sees results of a computer tomography or magnetic and resonance tomography, etc. Applications can carry also training character, prompting that else the patient needs to ask or appoint, in dependences on its complaints and diagnosis.

Other class of applications of augmented reality is an assembly, technical service, repair of the difficult technique. So in case of breakage in the cold winter on to the deserted road any woman can make primary car repairs. It puts on augmented reality glasses, via them carries out diagnostics, i.e. it looks on some detail of the car and it information is displayed. as has to look in a normal duty this detail as to check it also in a case to repair malfunctions. Similar applications are very popular also at offices, for example when replacing paper or a cartridge in the printer. When targeting on it of points, the user sees the short instruction from what party what to press and what it will lead to. The following class unites applications which add information to objects of various character, sometimes, this class is called by "the summary and visualization" [6, 8]. For example, in library, passing on ranks, to you information is displayed about the books standing on shelves, thereby, there is no need last or will bend that to get each of them. Other application helps teachers and teachers to trace progress of students. So looking at the student, through points of added realities, the teacher sees whether this student is allowed to examination, whether it is frequent it attended classes what GPA of progress, whether that the student came to examination, what ticket it extended as he spends at University, personal much time achievements, victories at competitions of projects, etc. Similar applications are popular in construction, architecture and design of rooms. Users can in actual time to add or clean the virtual object the existing space, to estimate dimensions, planning, spaciousness, color schemes, etc. To estimate the projected skyscraper will how successfully fit into the built-up area, what look will open from windows. Control of robots and unmanned aerial vehicles. Management of technique on distance is the complex challenge, there are problems with a transfer delay the operating information, reaction rate of the robot, distortion of information at to transfer [9, 10]. Therefore there was an idea to operate not directly, and to control the

virtual version of the robot. The user plans and defines operations of the robot with the help of a manipulation the local virtual version of the robot in actual scale time. Results are displayed immediately on actual object. After that, as the plan is checked and is chosen suitable, the user gives command to actual to the robot to implement the specified plan. These applications allow to avoid delays when performing sequence of actions and not to interrupt process, also to predict results of the taken actions on environmental objects, that most to prevent undesirable consequences.

Games and entertainments one of the most inventive classes of applications augmented reality [7]. So games on searching of things or tracings of participants games in the actual world can be supplied with three-dimensional graphics and the virtual hints, I make game unpredictable and fascinating. Also, having, for example the image, poster or photo of the player in soccer, the application distinguishes who it and provides you everything the available information on this player, the best goals personal achievements for which clubs are visualized by means of three-dimensional graphics, this athlete and many other things played. Of course theatrical sceneries, color decisions, lighting receptions can change and be imposed on alive acting, expanding and deepening space. Applications of augmented reality for military stand a separate class industries. Helmets of augmented reality are for a long time developed, which are actively used in simulators of flights by military planes various models in a wide criteria range and tasks. In investigation devices are used with support of AR which can impose drawings or a view from the satellite or information from the unmanned aerial vehicle immediately in the field of vision soldier. Apparently from the above, there is a lot of applications of augmented reality, they are very different and set contradictory tasks for telecom operators [11, 12, 13].

The greatest interest and popularity at users bought applications from a class "summary and visualization" and "assembly, maintenance and repair of equipment", therefore for model building of a traffic, model of interaction of separate elements, definitions of the main indexes of functioning of data of systems, were chosen applications of these groups.

3. Service model

According to the majority of definitions, providing service augmented reality consists in introduction in the field of perception person of padding information. Generally, appointment to facilitate the solution of various tasks, the bound to perception, the analysis and management. In fig. 1 the example of AR of the star chart is given (StarChart iPhone). Difference from just card or an interactive map consists in that, that its display is bound to position of the device in space (its targeting in a particular point of a palate), i.e. it is added with results information processing of the user.

In more general case process of providing service DR it is possible to consider as interactive interaction of the user with the applied functions



Figure 1. An example of augmented reality when looking at the starry sky (StarChart iPhone)

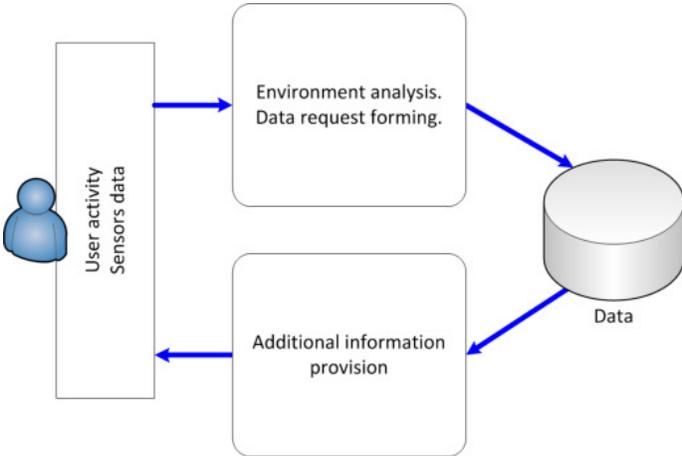


Figure 2. User interaction with the application function

realizing the analysis of a condition of its environment and providing it padding information, fig. 2.

As the purpose of AR is a granting to the user padding context-sensitive information, that quality of its granting it is necessary to consider from the

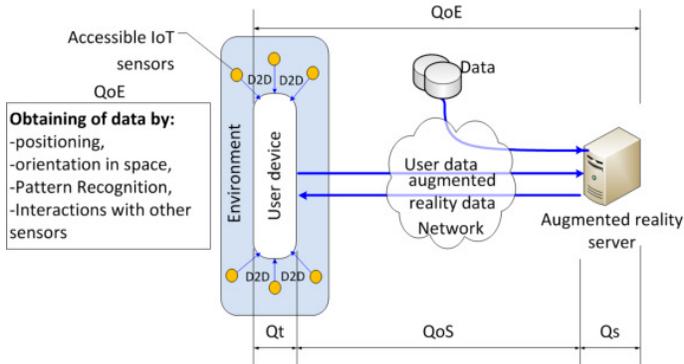


Figure 3. Interaction of the main elements in the provision of AR services

point of view of extent of achievement of this purpose that it is possible to characterize:

- degree of compliance to the provided padding information to needs of the user (compliance to purpose, volume, specification, etc.);
- degree of a susceptibility of the provided data (video this, graphics, sound, tables, the text and other elements of the interface with user, quality of their representation);
- timeliness of providing padding information.

Realization of service AR can be various, depending on the used technical means. In a set of these tools surely the subscriber device which can represent enters the smartphone, the tablet personal computer, multimedia glasses, a helmet, etc., allowing to organize the user interface, data acquisition about an environment and having rather high efficiency of computing devices. In particular, all functionality of AR can be realized as a part of the client application of such device. However, in many applications, for example, it is inexpedient or it is impossible to store all padding information and also to carry out all data processing user resources of the mobile device. Therefore, following the AR element is infocommunication making (communication network), which provides delivery of padding information to the user, databases and perhaps servers the carrying-out part of functions on processing information of the user, fig. 3. At existence in the user's environment the sensors of the Internet of Things (IoT) capable to provide the useful information, the D2D technology providing can be used direct connection of the subscriber device with them.

The basic elements providing quality of perception services by the user are the device of the user of AR, the service server AR, databases and communication network. All elements interact through a network communications (in the presence of IoT devices also D2D). In this system the main problem of ensuring quality of service is distribution functionality and

data between the client application of the device user, server of service and databases. This distribution it is reflected in a run time of functions on data processing, time deliveries of data through a communication network and on the traffic made in a network. Degree of compliance of padding information to requirements the user and degree of its susceptibility are defined by the organization services, existence of the required information and the organization of user interface. Service AR is interactive therefore timeliness providing information is one of the most important factors, defining its quality. Timeliness is characterized by time between an event the bound to changes of a condition of an environment or the user and the event characterizing availability to the user padding information. This time (delay) is defined by a row components, such as:

- data acquisition time about an environment (poll of sensors of a state, video, etc.) and their processings;
- time of delivery of data for the service server (if it is necessary);
- data processing time service server;
- time of delivery of data to the user;
- data presentation time.

At the organization of service with participation of the server data are requested and transferred to the user at change of a condition of an environment (change of an environment in the field of vision/perception of the user). Depending on functionality of service it can occur at change of position in space of the user (terminal) or some objects that is equivalent to change of a set of objects in the field of perception which require providing padding data. Identification of change can be made, for example, on the basis of the analysis of data on coordinates of the device and its orientation in space, discernments of objects by the analysis of video of data, etc.

4. Traffic Pattern

For the description of the traffic made by service it is necessary to connect volume given by the user and to the user of data at changes of his environment.

- service space model;
- model of an environment of the user;
- behavior model;

We will understand informational model of physical three-dimensional space in which there can be a user of service as space of service. The informational model includes the description of some objects which are in this space $X = \{\bar{x}_1, \bar{x}_2 \dots, \bar{x}_n\}$, where n total number of objects. The model of an environment of the user is a subspace service spaces, i.e. part of space restricted by opportunities perceptions (model of these opportunities). Environment, as a rule, it is attached to position of the user in space of service and includes in a set of objects $X^{(U)} = \{\bar{x}_1^{(U)}, \bar{x}_2^{(U)} \dots, \bar{x}_k^{(U)}\}$, where k number of the objects which are to the area perceptions of the user. The behavior model describes changes of position of the user and

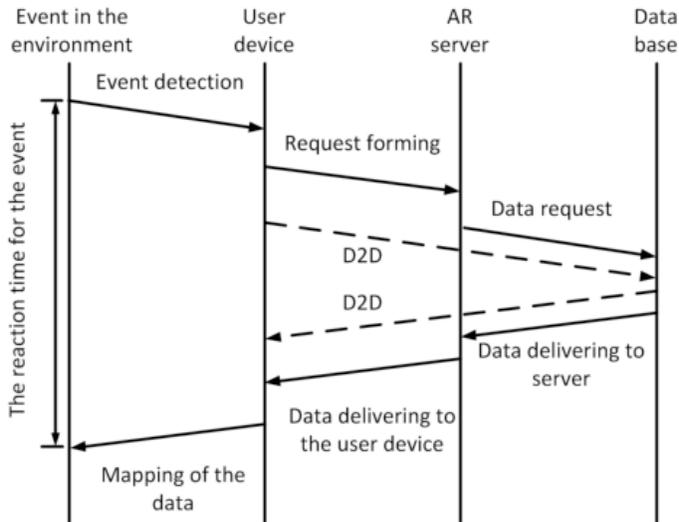


Figure 4. Data exchange diagram when providing AR services

it environments in service space. Changes in the user's environment can occur as owing to movements of the user, and movements of objects in service space. The change caused emergence in an environment of the user of new object \bar{x}_i , leads to inquiry data on this object.

The algorithm of realization of service has to provide realization following functions:

- identification of an event of change of an environment and calculation change parameters;
- information request about change of an environment;
- data acquisition and their display.

The possible chart of data exchange is given in fig. 4.

As an environment also as well as space of service is restricted (area of perception) physical three-dimensional space, changes in it can be described as a stream of events of the bound to emergence in it objects. Objects can enter through its borders owing to movement of these borders or objects. In case of the former, movement of borders is bound to behavior of the user, and in the second to behavior of objects. In that and other case quality of functioning of system will depend on its ability in due time to serve events of this stream. Thus, the problem of ensuring quality of service can be considered as the choice of parameters of system (efficiency, a channel capacity, distribution of its functionality) from the characteristic of a stream of events and load of the system made by this stream.

Properties of a stream of events substantially define properties of a data flow between elements of system. For example, in system of positioning on a district map such stream is defined by events of change of coordinates of the user and is defined by characteristics of his driving, in system it will decide by their orientation in space on use of points of DR and it is possible events the bound to data transmission about objects under review or their characteristics.

It is apparent that characteristics of a stream will depend on distribution and characteristics of objects in service space, and also characteristics of driving of the user. Let's make an assumption that objects in space of service are distributed in a random way (form the Poisson field) and are not mobile, only the user is mobile. Then, change of position of the user is equivalent to change of its environment. Taking into account properties of space of service and an environment, this change can be described by the volume or the area. Let's consider 2D option and we will describe the user's environment r radius circle, and we will consider the speed of movement of the user of v as a constant, Then that during t change of an environment will be defined by number of new objects in the area defining the user's environment.

Let's estimate amount of new objects in an environment during t as (1)

$$n(t) = \tilde{S}(L(t)) \rho \tag{1}$$

Where $\tilde{S}L(t)$ - the area of change of an environment; ρ - density of objects (objects/sq.m).

The model illustrating movement of the user and change of its environment is given in fig. 5. Shift of the circle representing the user's environment from the initial point on distance of L is led to formation of area in the form of a sickle (the shaded area) by which defines change of an environment. Objects it is in the field identified, according to algorithm of providing service therefore requests for providing padding information are formed.

From the given drawing, the area of the shaded area can be determined by a formula (2) the Number of new objects in the area can be defined as

$$\tilde{S}(L) = \pi r^2 - 2 \left(r^2 \arccos \left(\frac{L}{2r} \right) - \left(\frac{L}{2} \right) \sqrt{r^2 - \left(\frac{L}{2} \right)^2} \right) \tag{2}$$

The number of the new objects in the field can be defined as (3)

$$n(t) = \tilde{S}(L)\rho \tag{3}$$

where ρ -density of objects (objects/sq.m). Considering this process in dynamics, i.e. when driving the user the stream of events (requests of data) takes place. Intensity of a stream of events (requests for data) can

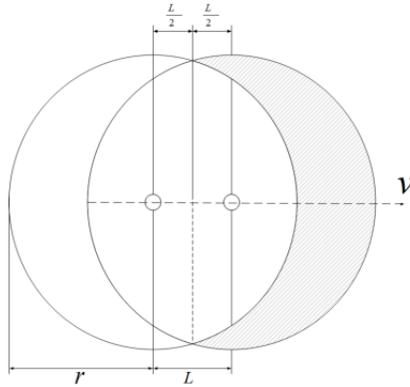


Figure 5. Movement of the user and change of its environment

be defined as number of objects in a small increment of the area of the considered figure

$$\lambda_r = \frac{d\tilde{S}(L)}{dL} \rho v \tag{4}$$

Where ρ - density of objects (objects/sq.m); v - speed of movement (m/s). Derivative of expression (2) in a point $L = 0$

$$\frac{d\tilde{S}(L)}{dL} = [L = 0] = 2r \tag{5}$$

Then taking into account (4) and (5) we got (6)

$$\lambda_r = 2r\rho v \tag{6}$$

Considering properties of the Poisson field accepted for model the amount of objects in some restricted area is casual, distributed under the law of Poisson and depends only on the area (or volume) the considered area. Therefore, for the accepted model the stream of inquiries will represent the elementary stream for which the probability of the interval of time t of receipt of requests k will be defined as (7)

$$p_k = \frac{(\lambda_r t)^k}{k!} e^{-\lambda_r t} = \frac{(2r\rho v t)^k}{k!} e^{-2r\rho v t} \tag{7}$$

The traffic stream made as a result of providing service is defined by a stream of replies to the requests of data. Generally, the simple package of data, and a stream of packages (transfer of video or audio of data) can be the answer both. Intensity of this stream can be described as (8)

$$\lambda_r = \lambda_s \eta \quad (8)$$

Where η - the average number of the packages necessary for realization of inquiry.

By transfer of video data can exceed intensity of inquiries in tens and hundreds of times. Taking into account requirements to quality it leads to essential growth of requirements to a channel capacity of a communication network. The physical sizes of an environment of the user are, as a rule, commensurable with a radius of communication, the wireless technologies used for the PAN organization, for example family of the WiFi standards. Many objects of services AR (elements of city infrastructure, vehicles, household appliances) can be equipped with clusters of access and necessary data which it can be provided to users. Therefore, use of the D2D technologies can be the possible decision providing essential decrease in a traffic on a communication network. In that case the traffic of data can be delivered immediately from object to the terminal of the user of fig. 4 (dashed line). Intensity of this stream will be defined as (9)

$$\lambda_r = \lambda_s \eta_{D2D} \quad (9)$$

Where $\eta_{D2D} = (1 - \gamma)\eta$, γ - a share of the objects of an environment supporting the D2D technology.

Certainly, application of D2D technologies is possible only when objects of service are the physical objects mentioned above which I can be equipped with the corresponding communication centers. Rather wide range of services nevertheless demands interaction with remote databases and problem solving of ensuring quality.

5. Conclusions

1. Rendering of services of AR demands differentiated approach to definition of their indexes of quality depending on purpose of service.

2. Quality of providing service DR is characterized by degree of compliance to purpose, degree of a susceptibility of the provided data and timeliness of providing information.

3. Properties of the traffic made by service are defined by behavior of the user, an environment of the user and way of data presentation.

4. Intensity of a traffic of services AR when using multimedia data it is comparable to a traffic of services of transfer of video that taking into account interactivity of service and requirements to quality leads to increase in requirements to a channel capacity of a communication network.

5. One of ways of reduction of requirements to a channel capacity of a network can be application of the D2D technologies which use leads to short circuit of a traffic, between the terminal of the user and objects of an environment, passing networks of telecom operators.

Acknowledgments

The reported study was supported by grants from the RFBR, research project No. 16-37-00209 mol a “Development of the principles of integration the Real Sense technology and Internet of Things”.

References

1. *Federico Bergenti, Danilo Gotta*. «Augmented Reality for Field Maintenance of Large Telecommunication Networks», Conference and Exhibition of the European Association of Virtual and Augmented Reality, 2014.
2. *Sorensen L., Skouby K.E.* Use scenarios 2020 – a worldwide wireless future. Visions and research directions for the Wireless World / Outlook. Wireless World Research Forum. – , N4, July 2009.
3. *Ganapati, Priya*. "How it Works: Augmented Reality." Wired. Aug. 25, 2009. [http : //www.wired.com/gadgetlab/2009/08/total-immersion/](http://www.wired.com/gadgetlab/2009/08/total-immersion/)
4. *Rocha, Roberto*. "Tech's new reality." Montreal Gazette. Oct. 23, 2009. [http : //www.montrealgazette.com/technology/Tech + reality/2138659/story.html](http://www.montrealgazette.com/technology/Tech+reality/2138659/story.html)
5. *Wortham, Jenna*. "UrbanSpoon Makes It Easier to 'Scope' Out Restaurants." New York Times. Oct. 10, 2009. [http : //bits.blogs.nytimes.com/2009/10/14/urbanspoon - makes - it - easier - to - scope - out - restaurants/](http://bits.blogs.nytimes.com/2009/10/14/urbanspoon-makes-it-easier-to-scope-out-restaurants/)
6. *Ensha, Azadeh*. "Another Augmented-Reality App for the iPhone." New York Times. Oct. 15, 2009. [http : //gadgetwise.blogs.nytimes.com/2009/10/15/augmented - reality - apps - continue - to - roll - out/](http://gadgetwise.blogs.nytimes.com/2009/10/15/augmented-reality-apps-continue-to-roll-out/)
7. *Inbar Ori*. "Top 10 reality demos that will revolutionize video games." Games Alfresco. March 3, 2008. [http : //gamesalfresco.com/2008/03/03/top - 10 - augmented - reality - demos - that - will - revolutionize - video - games/](http://gamesalfresco.com/2008/03/03/top-10-augmented-reality-demos-that-will-revolutionize-video-games/)
8. *Parr, Ben*. "Easter Egg: Yelp Is the iPhone's First Augmented Reality App." Mashable. Aug. 27, 2009. [http : //mashable.com/2009/08/27/yelp - augmented - reality](http://mashable.com/2009/08/27/yelp-augmented-reality)
9. *Koucheryavy A., Vladyyko A., Kirichek R.* State of the art and research challenges for public flying ubiquitous sensor networks. Lecture Notes in Computer Science, 2015, V.9247. PP. 299-308.
10. *Koucheryavy A., Paramonov A.* Analysis of routes in the network based on a swarm of UAVS. Dao N., Koucheryavy A., Paramonov A. Lecture Notes in Electrical Engineering. 2016. V. 376. PP. 1261-1271.
11. *Paramonov, A., Koucheryavy, A.* M2M traffic models and flow types in case of mass event detection (2014) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8638 LNCS, pp. 294-300.

12. *M.Z.Shafiq and all.* A First Look at Cellular Machine-to-Machine Traffic: Large Scale Measurement and Characterization. 12th ACM Sigmetrics Performance International Conference. June 11-15, London, England, UK, 2012.
13. *Kirichek R., Vladyko A., Zakharov M., Koucheryavy A.* Model networks for Internet of Things and SDN. Kirichek R., Vladyko A., Zakharov M., Koucheryavy A. 18th International Conference on Advanced Communication Technology (ICACT) 2016. PP. 76-79.

UDC 004.4

The three-dimensional simulation framework for interference and SIR assessment

R. Kovalchukov*, A. Samuylov*[†], D. Moltchanov*[†],
S. Andreev[†], K. Samouylov*

* *Department of Applied Probability and Informatics
RUDN University*

Miklukho-Maklaya st 6, Moscow, 117198, Russia

[†] *Department of Electronics and Communications Engineering
Tampere University of Technology*

Korkeakoulunkatu 10, Tampere, 33720, Finland

Abstract. Within the framework of stochastic geometry, the performance of modern wireless networks is conventionally assessed using two-dimensional models. However, as the wireless systems evolve further, the role of the third dimension may start to play an important role. The underlying reasons are the projected use of highly directional antennas and different heights of communicating entities. In this work, we build a simulation tool for interference and SIR characterization in three-dimensional network deployments.

Keywords: interference, signal-to-interference ratio, blocking, directional antennas, millimeter waves, three-dimensional model.

1. Introduction

The performance of wireless networks has been conventionally assessed using the tools of stochastic geometry [1, 2]. Most of the models proposed so far for different wireless network deployments including 4G and pre-5G systems are based on two-dimensional abstractions, e.g., [3, 4]. The rationale is that the height of transmitters and receivers does not affect the performance metrics of interest. However, as the directivity of antennas increases with the use of extremely high frequency (EHF), this assumption may no longer hold.

In this work, taking into account properties of millimeter wave radiation patterns and propagation including both directivity and absorption, we develop a three-dimensional simulation framework for interference and SIR assessment in dense device-to-device (D2D) communications deployments with different heights of transmitters and receivers.

2. System Model

2.1. Deployment model

In this work, we study two scenarios shown in Fig. 1. The considered network consists of communicating pairs of devices, transmitters and receivers. Each pair can belong to one of three types: tagged pair – colored

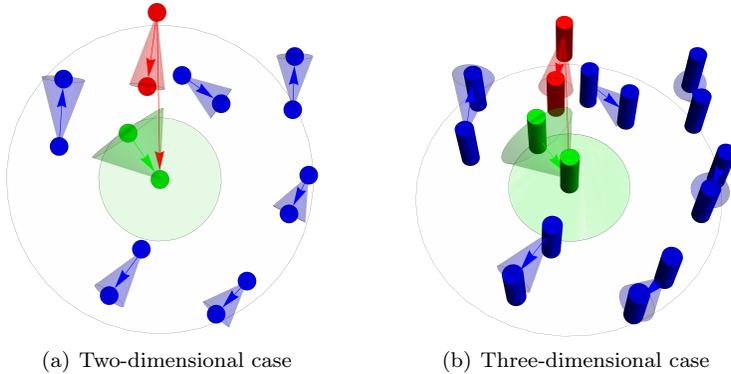


Figure 1. The considered two- and three-dimensional D2D deployments.

green in Fig. 1, interfering pair – colored red and non-interfering pairs – colored blue. We first randomly select a point $P \in \mathbb{R}^2$ and associate it with the tagged receiver. The transmitters are deployed as a hard-core point process with intensity λ and parameter r_B within a circular area of interest of radius T_R , $T_R < I_R$, for the tagged transmitter and I_R for all other transmitters around the tagged receiver. Here, T_R is the maximum distance of possible transmission and the value of I_R is chosen such that the interference from devices that are further away from receiver than I_R is negligible. Receivers associated with transmitters are distributed uniformly inside the circle of radius T_R , around the respective transmitters.

All of the transmitters are assumed to use the same transmit power, operate at the same frequency, and create interference to the receivers. In two-dimensional case, we completely disregard the heights of all objects deployed in the scenario. In the three-dimensional scenario, additional dimension is used to represent blockers' and devices' heights h_B and h respectively. That is, bodies of users carrying communicating devices are represented as cylinders with random heights h_B . According to [5], the users' height is assumed to follow normal distribution with mean and standard deviation, x and y , respectively, while their radii are constant and equal to r_B . Each human is associated with a device whose height is uniformly distributed over $(0, h_B)$. For both scenarios, we are interested in the aggregate interference and SIR at the tagged receiver.

2.2. Antenna radiation pattern models

Following [6], we approximate the directive antenna radiation pattern by the so-called cone model. Accordingly, a cone vertex is located at the antenna, meaning that the strength of the radiating radio waves outside

the cone is negligible. This model is an abstraction assuming that isotropic receive level depends only on the separation distance from the transmitter.

In the considered scenario, every transmitter is equipped with a directional antenna having a cone-shape wavefront (blue cones in Fig. 1). Receive antennas are also directional (green cone in Fig. 1) and have sensitivity only in a certain direction. This implies that any transmitter and receiver pair must be aligned with respect to each other.

Parameterization of the antenna model is provided by amplification coefficient G_A corresponding to a directivity angle α of the antenna. The latter can be found using the surface area of a wavefront $S_A = \frac{P_{Tx}}{S_{Rx}(f,r)} = 2\pi rh$, where P_{Tx} is the total transmit power, $S_{Rx}(f,r)$ is the power spectral density (psd) at the wavefront, r is the separation distance, f is the operational frequency, $h = r[1 - \cos(\alpha/2)]$, and α is the directivity angle.

According to free-space path loss model, we have

$$S_{Rx}(f,r) = \frac{P_{Tx}}{S_A} = \frac{G_A}{4\pi f^2} r^{-2} P_{Tx},$$

implying that the antenna gain $G_A = \frac{2}{1 - \cos(\alpha/2)}$.

2.3. Propagation and blockage models

A distinguishing feature of the EHF channels is the presence of the molecular absorption losses making wideband wireless channels extremely frequency selective [7]. In EHF band, especially in unlicensed 60GHz band, these losses are caused by absorption lines of O_2 molecules.

The received psd in the EHF band can be written as

$$S_{Rx}(f,r) = \frac{S_{Tx}(f)}{L_A(f,r)L_P(f,r)},$$

where $S_{Tx}(f)$ stands for the transmit signal psd, $L_A(f,r)$ represents the absorption loss, and $L_P(f,r)$ is the propagation losses.

The absorption loss is $L_A(f,d) = \frac{1}{\tau(f,r)}$, where $\tau(f,r)$ is the transmittance of the medium following the Beer-Lambert law, $\tau(f,r) \approx e^{-K(f)r}$, where $K(f)$ is the overall absorption coefficient of the medium available, for instance, in HITRAN database [8].

Finally, to capture the effect of blockage, we assume that the presence of any human along the line-of-sight (LoS) path connecting any transmitter and receiver pair completely blocks the transmission.

3. Conclusions

In this work, we constructed a simulation framework for interference and SIR assessment in three-dimensional scenarios taking into account the

specifics of millimeter wave frequency band including the use of highly directional antennas, propagation, and blocking. In our future studies, we will rely on this model to identify the cases when the use of conventional two-dimensional scenarios is insufficient for accurate performance characterization of future millimeter wave communications systems.

Acknowledgments

The reported study was funded within the Agreement No. 02.a03.21.0008 dated 24.04.2016 between the Ministry of Education and Science of the Russian Federation and RUDN University. The reported study was partially supported by the RFBR, research projects No. 15-07-03051, 15-07-03608, 16-07-00766.

References

1. F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks, Part I: Theory*. Now Publishers, 2009.
2. M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.
3. J. Andrews, R. Ganti, M. Haenggi, N. Jindal, and S. Weber, “A primer on spatial modeling and analysis in wireless networks,” *Communications Magazine, IEEE*, vol. 48, pp. 156–163, Nov. 2010.
4. V. Petrov, M. Komarov, D. Moltchanov, J.-M. Jornet, and Y. Koucheryavy, “Interference analysis of ehf/thf communications systems with blocking and directional antennas,” in *Proc. of IEEE Globecom*, Dec. 2016.
5. C. L. Ogden, C. D. Fryar, M. D. Carroll, and K. M. Flegal, “Mean body weight, height, and body mass index, united states 1960–2002,” *Centers for Disease control and prevention*, October 2004.
6. S. Singh, R. Mudumbai, and U. Madhow, “Interference analysis for highly directional 60-ghz mesh networks: The case for rethinking medium access control,” *IEEE/ACM Transactions on Networking (TON)*, vol. 19, no. 5, pp. 1513–1527, 2011.
7. J. M. Jornet and I. F. Akyildiz, “Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band,” *IEEE Trans. Wir. Comm.*, vol. 10, pp. 3211–3221, Oct. 2011.
8. www.cfa.harvard.edu, “Hitran: High-resolution transmission molecular absorption database,” tech. rep., Harvard-Smithson Center for Astrophysics, 2014.

УДК 004.415.538

Методы комплексного тестирования устройств Интернета Вещей

В. А. Кулик, А. И. Выборнова

*Кафедра сетей связи и передачи данных
Санкт-Петербургский государственный университет телекоммуникаций
им. М.А.Бонч-Бруевича
пр. Большевиков, д.22, к.1, Санкт-Петербург, Россия, 193232*

Аннотация. В данной статье будет освещен ряд проблем связанных с проведением комплексного тестирования устройств Интернета Вещей. В частности будут рассмотрены проблемы связанные с нагрузочным тестированием, тестированием безопасности и удаленным тестированием. Для каждого из этих видов тестирования будет предложена модель, описывающая процесс тестирования.

Ключевые слова: Интернет Вещей, интернет вещь, нагрузочное тестирование, безопасность, облако, анализ трафика.

1. Введение

В настоящее время, в связи с интенсивным внедрением концепции Интернета Вещей (ИВ) в современную сеть связи общего пользования (ССОП), число устройств, подключенных к сети Интернет, продолжает увеличиваться. Согласно прогнозам, число устройств в ССОП к 2020 году преодолет границу в 50 млрд. устройств [1]. Такое огромное количество устройств, функционирующих в сети Интернет в первую очередь связано с концепцией всепроникающих сенсорных сетей, которая развивается в рамках отрасли Интернета Вещей. Данная концепция предполагает объединения большого числа устройств, являющемся совокупностью разных видов устройств беспроводной связи и различными типами датчиков, в беспроводную сеть, посредством ССОП [2,3]. Предполагается, что данные устройства будут охватывать большое число отраслей, например медицину, контроль правопорядка, промышленность, экологический мониторинг, военное дело и др. Подразумевается, что столь огромное число устройств ИВ изменит сам характер трафика ССОП [4, 5]. В связи с этим, поднимается ряд вопросов, связанных с устойчивостью современной сети связи к трафику ИВ и защищенности конфиденциальных пользовательских данных.

На настоящий момент не существует единых методов тестирования устройств Интернета Вещей в ССОП [6]. В данной статье предлагаются ряд методов для тестирования данных устройств, в частности методы тестирования оборудования ССОП (маршрутизаторы, коммутаторы) на устойчивость к трафику ИВ, тестирования безопасности устройств ИВ, с помощью анализа сетевого трафика и удаленного тестирования устройств ИВ.

2. Нагрузочное тестирование

В настоящий момент, в связи с переходом от сетей связи NGN к сетям связи пост-NGN и наблюдающимся в связи с этим рост числа устройств в сети Интернет, возникает проблема тестирования сетевого оборудования к трафику Интернета Вещей. Сам трафик ИВ, генерируемый устройством ИВ, в первую очередь характеризуется [4, 5]:

- 1). Более высокой долей передаваемого трафика, чем запрашиваемого;
- 2). Превалированием размера заголовка пакета, над размером поля данных;
- 3). Высокой частотой отправки пакетов;
- 4). Иными законами временного распределения пакетов. Подразумевается, что помимо того что трафик ИВ может является самоподобным трафиком, он еще может являться антиперсистентным.

Данные характеристики подтверждают необходимость тестирования сетевого оборудования на устойчивость к трафику ИВ.

Для реализации устройства нагрузочного тестирования предлагается модель изображенная на рис. 1. Согласно данной модели устройство тестирования должно обладать следующими функциями:

- 1). Параллельная генерация трафика от множества виртуальных устройств ИВ. Т.е. возможность создания на устройстве тестирования множества виртуальных устройств ИВ, имеющих уникальные параметры (уникальный MAC, IP, идентификационный номер и т. д.);
- 2). Виртуальное устройство ИВ, имеет один из основных типов устройств ИВ [7], например:

- сенсор (устройство преобразующее и отправляющее данные, получаемые от датчиков. Датчик — это электронное устройство, которое измеряет физическое состояние или химический состав и доставляет электронный сигнал, соответствующий наблюдаемой характеристике [7, 8]);

- актуатор (устройство, которое инициирует физическое действие после возбуждения входным сигналом [7, 8]);

- мультимедиа устройство (устройство отправляющее мультимедийную информацию. Мультимедийная информация – это цифровая информация, в которой для информирования или развлечения пользователей используется много видов информационного контента и обработки информации, например, текст, изображения, аудио, видео, трехмерные панорамные изображения и цифровые карты [7, 8]);

- 3). Поддержка основных сценариев работы ИВ, таких как: - Периодичная отправка данных. Каждое устройство периодически отправляет данные на удаленный облачный сервер; - Отправка данных по запросу; - Воздействие на окружающий мир устройством, по управляющему сигналу.

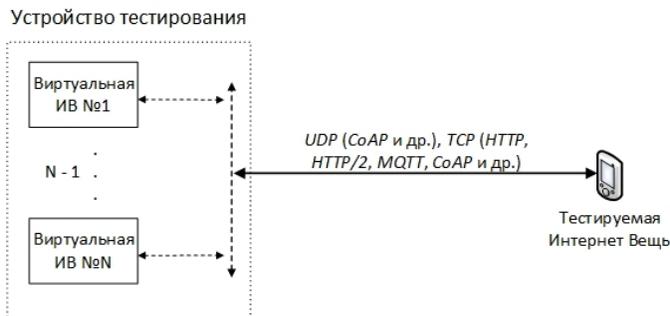


Рис. 1. Модель нагрузочного тестирования трафиком Интернета Вещей

4). Возможность генерации различных типов трафика от каждого виртуального устройства Интернета Вещей по определенному закону временного распределения [4, 5].

5). Связь с удаленным тестовым сервером, необходимым для тестирования сценариев, в которых необходимо использование протоколов транспортного уровня с поддержкой соединения. Также данный сервер, необходим для измерения уровня задержки при высоких уровнях нагрузки оборудования трафиком Интернета Вещей.

Важным моментом, помимо поддержки возможностей вышеуказанной модели нагрузочного тестирования, является поддержка устройством тестирования всех основных типов физических интерфейсов передачи данных, в частности интерфейс PON, интерфейс Ethernet, интерфейс IEEE 802.11.

Сам процесс нагрузочного тестирования напрямую зависит от мощности самого тестируемого оборудования и пропускной способности канала связи. В случае маломощных устройств тестирование проходит путем загрузки канала связи большим числом пакетов малого размера, каждый из которых имеет свои уникальные параметры. В случае более мощного сетевого оборудования, на УТ необходимо регистрировать множество виртуальных устройств ИВ на тестируемом оборудовании, согласно стандарту IEEE 802.1x [9] и затем генерировать от каждого из ВУ пакеты ИВ.

3. Тестирование безопасности

В связи с большим количеством устройств ИВ, подключенных к сети Интернет и отсутствием общепринятых методов обеспечения криптографической и сетевой безопасности для устройств ИВ, появляется значительное количество угроз для конфиденциальной пользовательской информации. Данные угрозы создают потребность в тестировании Интернет Вещей на наличие уязвимостей сетевой безопасности.

Одной из распространенных уязвимостей устройств Интернета Вещей является перехват конфиденциальных данных, идущих от устройства Интернета Вещей в ССОП. В связи с этим предлагается проводить тестирование Интернет Вещей на уязвимость сетевой безопасности методом перехвата и анализа сетевого трафика. Метод заключается в подключении к каналу связи, по которому идет трафик Интернета Вещей, устройства перехвата и анализа трафика ИВ [10]. В ряде случаев, если не используются алгоритмы шифрования, возможно проанализировать информацию, находящейся в поле данных перехваченных пакетов. Далее если устройству тестирования удастся определить, используемый устройством ИВ протокол передачи данных прикладного уровня, то проводится анализ поля данных данного пакета для выделения ключевых полей. Например, в протоколе MQTT присутствует поле topic, в котором находится информация о брокере на который подписано устройство, т. е. фактически — это адрес приложения, которое отвечает за взаимодействие устройства ИВ и MQTT-сервера [11]. После проведенного анализа устройство тестирования на основе полученных данных определяет тип устройства Интернета Вещей.

Таким образом, после определения формата сообщения, возможна фальсификация данных, получаемых сервером от устройства ИВ, путем отправки «ложного» пакета, дублирующего перехваченные от устройства ИВ пакеты, но с измененным содержанием, на удаленный облачный сервер, связанный с атакуемым устройством ИВ. В случае успешного проведения тестирования устройство ИВ считается уязвимой для атаки, связанной с перехватом пользовательской информации.

4. Удаленное тестирование

На данный момент существует проблема невозможности проведения тестирования для уже подключенных и функционирующих в ССОП, устройств ИВ. Данный процесс должен выполняться без участия специализированных кадров и является полностью автоматизированным для удешевления процесса тестирования.

Для решения проблем удаленного тестирования устройств ИВ предлагается модель (рис. 2) состоящая из:

- 1). Удаленного сервера тестирования (Cloud testing server). Сервер отвечающий за контроль процесса тестирования, путем отправки управляющих команд устройству тестирования. Также сервер проводит сбор и анализ статистики, получаемой от устройства тестирования;

- 2). Маршрутизатор (Switch). Отвечает за создания локального сегмента ССОП, в котором функционируют устройство ИВ и устройство тестирования;

- 3). Устройство тестирования (Testing device). Данное устройство находится в одном адресном пространстве с тестируемым устройством

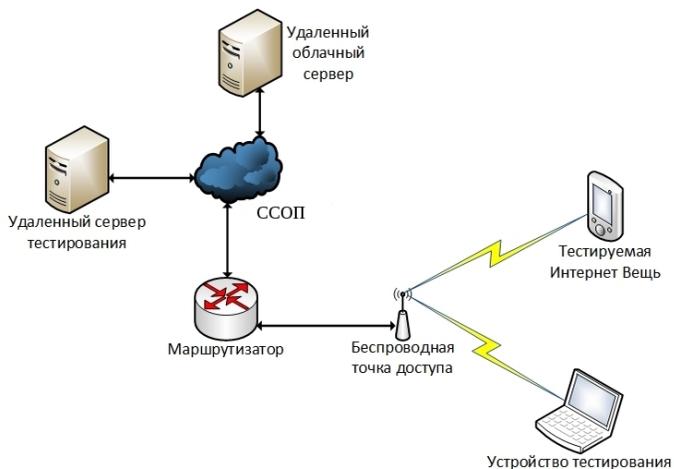


Рис. 2. Пример модели удаленного тестирования с доступом к беспроводному каналу связи

Интернета Вещей и имеет доступ к каналу связи, по которому устройство ИВ транслирует данные в ССОП;

4). Удаленный облачный сервер (Cloud server). Сервер, с которым происходит взаимодействие устройства Интернет Вещей;

5). Тестируемое устройство Интернет Вещей (Internet of Thing device under testing). Устройство ИВ, находящейся в одном адресном пространстве с устройством тестирования и взаимодействующее с удаленным облачным сервером.

Такой программно-аппаратный комплекс возможно организовать на основе модельной сети лаборатории Интернета Вещей СПбГУТ им. М.А.Бонч-Бруевича [12].

Удаленное тестирование устройства Интернета Вещей предлагается проводить по следующему алгоритму (рис. 3):

1). Тестируемая интернет вещь и устройство тестирования подключаются к одному маршрутизатору (локальной сети);

2). На устройство тестирования, в фоновом режиме, запускается программное обеспечение, отвечающее за взаимодействие с удаленным сервером тестирования;

3). Устройство тестирования предоставляется доступа к каналу связи, по которому передаются данные от устройства ИВ;

4). Пользователь, через интерфейс взаимодействия с удаленным сервером тестирования, выбирает вид тестирования Интернет Вещи;

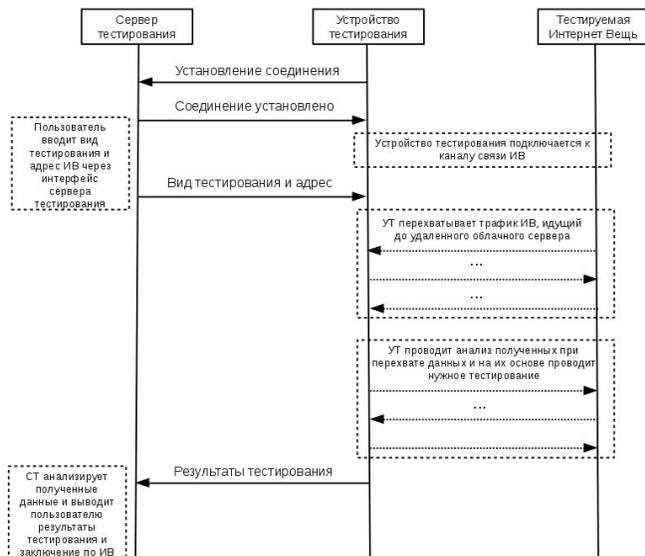


Рис. 3. Модель взаимодействия устройств, при проведении удаленного тестирования

5). Облачный сервер отправляет управляющие сигналы на устройство тестирования. Устройство тестирования переключается в режим ожидания;

6). Тестируемое устройство Интернета Вещей активируется и начинает взаимодействие с удаленным облачным сервером, по заранее прописанному в устройстве сценарию;

7). Устройство тестирования перехватывает трафик, идущий от устройства Интернет Вещей, и проводит его анализ. Фильтр для перехвата трафика определяется, с помощью заранее указанного пользователем, с помощью интерфейса взаимодействия, IP-адреса устройства;

8). Устройство тестирования, по определенным критериям (например, коэффициент Хёрста, протокол взаимодействия транспортного уровня, порт взаимодействия и др.) определяет тип устройства ИВ и протокол взаимодействия с облачным сервером;

9). Далее проводится одно из возможных видов тестирования [13], в зависимости от того какой вид был выбран пользователем в интерфейсе взаимодействия с удаленным сервером;

10). Устройство тестирования отправляет статистику о результатах тестирования на удаленный сервер тестирования;

11). Сервер тестирования, на основе анализа полученных данных, выдает пользователю заключение о устройстве Интернета Вещей, по выбранному виду тестирования.

5. Заключение

В данной статье были определены проблемы комплексного тестирования устройств ИВ и предложен ряд моделей для их решения. В частности были разработаны модели нагрузочного тестирования сетевого оборудования на устойчивость к трафику ИВ, тестирования безопасности устройства ИВ, удаленного тестирования устройств ИВ.

Данные методы закрывают ряд проблем, связанных с комплексным тестированием устройств Интернета Вещей и в дальнейшем могут быть реализованы в программно-аппаратном комплексе удаленного тестирования устройств Интернета Вещей, используемом для тестирования узлов наземного сегмента летающей сенсорной сети.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта N15-07-09431a "Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей".

Литература

1. *Dave Evans* The Internet of Things How the Next Evolution of the Internet Is Changing Everything // CISCO White Papers // April 2011.
2. *А.Е.Кучерявый* Интернет Вещей // Электросвязь. - 2013. - № 1. с.21-24.
3. *А.Е.Кучерявый, А.Е.Проконьев, Е.А.Кучерявый* Самоорганизующиеся сети. – СПб.: Любавич, 2011. – 312 с.
4. *Anastasia Vybornova, Andrey Koucheryavy* Traffic Analysis in Target Tracking Ubiquitous Sensor Networks // 14th International Conference, NEW2AN 2014 and 7th Conference, ruSMART 2014 vol. 8638 // Springer International Publishing // August 27-29, 2014.
5. *Парамонов А.И.* Разработка и исследование комплекса моделей трафика для сетей связи общего пользования // Автореферат, СПбГУТ // Санкт-Петербург 2014 год.
6. *Васильев А.Б., Тарасов Д.В., Андреев Д.В., Кучерявый А.Е.* Тестирование сетей связи следующего поколения. М.:ФГУП ЦНИИС, 2008 г. – 144 с.
7. *Peter Waher* Learning Internet of Things – Birmingham, UK, Packt Publishing Ltd. // Jan 2015.
8. *Y.2069.* Terms and definitions for the Internet of things // Recommendation ITU-T // Jul 2012.

9. IEEE 802.1x-2010 Port-Based Network Access Control // IEEE Standard for Local and metropolitan area networks // IEEE Computer Society, Feb 2010.
10. *Киричек Р.В., Кулик В.А., Владыко А.Г., Богданов И.А., Кучерявый А.Е.* Ложные облака для Интернета Вещей. Методы защиты // Информационные технологии и телекоммуникации. — 2015. — № 3. — С. 27–39.
11. MQTT Version 3.1.1 // OASIS Standard // OASIS Open 2014.
12. *Kirichek R., Koucheryavy A.* Internet of Things Laboratory Test Bed // Lecture Notes in Electrical Engineering. 2016. Vol. 348. PP. 485–494.
13. *Кулик В.А., Киричек Р.В., Бондарев А.Н.* Методы исследования беспроводных каналов связи Интернета Вещей в условиях совместной работы // Информационные технологии и телекоммуникации. - 2015. - № 1. - С. 106-114.

UDC 004.415.538

Methods of complex testing the devices of the Internet of Things

V. A. Kulik, A. I. Vybornova

*The Department of communication networks and data transmission
Saint-Petersburg State University of Telecommunications
Bolshevikov st., 22, building 1, St. Petersburg, Russia, 193232*

This article will cover a number of problems associated with complex testing devices of the Internet of Things. In particular, it will consider a number of problems associated with load testing, security testing and remote testing. For each of these types of test will suggest a model that describes the testing algorithm

Keywords: The Internet of Things, internet of thing device, benchmark, security, cloud, traffic analysis.

UDC 519.872

The total capacity of customers in the MMPP/GI/ ∞ queueing system

E. Lisovskaya*, S. Moiseeva*, M. Pagano[†]

* *Department of probability theory and mathematical statistics,
Tomsk State University*

Lenin ave. 36, Tomsk, 634050, Russia

[†] *Department of International Engineering,
University of Pisa
Via Caruso 16, Pisa, 56122, Italy*

Abstract. In the paper, the infinite-server queueing system with a random capacity of customers is considered. In this system, the total capacity of customers is analysed by means of the asymptotic analysis method with high-rate Markov Modulated Poisson Process arrivals. It is obtained that the stationary probability distribution of the total customer capacity can be approximated by the Gaussian distribution. Parameters of the approximation is also derived in the paper.

Keywords: infinite-server queueing system, customer with random capacity, Markovian Modulated Poisson Process.

1. Introduction

In the design of messages processing and transmission systems, determining the memory capacity required for information storage is a relevant open issue [1, 2]. The total capacity is a random quantity and in queueing theory it is given by the sum of the lengths of all messages, which are waiting in the buffer or currently processed by servers.

Since in real systems customers are heterogeneous (for instance in computer networks packet size may vary from a few tens of bytes to 1500 bytes in case of Ethernet links), this paper focuses on the analysis of queueing system (QS) with random customer capacity. The main classes of models used for such models and their applications in real information systems are given [1, 2].

There are several works on the study of such systems with Poisson arrival process and service time independent of the customer capacity. For example, in [3] for systems with limited total capacity the generalization of the Erlang problem is considered in stationary conditions. Moreover, in [4, 5], the stationary distribution of the customers number and the probability of losses are obtained for systems with limited memory. Finally, in [9], the authors consider systems with service time depended on the customers capacity or with waiting time restrictions, an assumption very relevant for real-time applications.

It is important to point out that most known results are obtained for queueing systems and networks with Poisson arrivals. Unfortunately, it

has been proved that the Poisson model is suitable only for few cases of modern telecommunication streams [6] and, in general, the correlation among arrivals must be taken into account. Therefore, many researches use more complex models of arrivals, such as Markovian Arrival Processes (MAPs) [7] or semi-Markov processes [8].

The main contribution of this paper consists in extending previous works on random capacity customers to the case of correlated arrivals. In more detail, the problem statement is formally defined in section 2 and in section 3 the corresponding Kolmogorov differential equations are derived. Then section 4 presents the results of the asymptotic analysis, focusing on first and second order approximations. Finally, the main findings are summarized in section 5.

2. Problem statement

In this paper, the MMPP/GI/ ∞ QS with random capacity customers is studied. The arrival process is a Markov Modulated Poisson Process (MMPP), a widely-used special case of MAP [7]. The system has an unlimited number of servers and service times on each server are i.i.d. with distribution function $B(x)$. All customers have a random capacity $\nu > 0$ with the probability distribution $G(y) = P\{\nu < y\}$ and the customers capacity are independent. Moreover, we assume that service time and customers capacity are mutually independent. After the service, customers leave the system and carry out the capacity.

Let us denote the number of customers in the system and the total customers capacity at time t by $i(t)$ and $V(t)$, respectively. We consider two-dimensional stochastic process $\{i(t), V(t)\}$, which is not Markovian. Therefore, we propose the dynamic screening method for its investigation [10].

For the screened process construction, we fix some point in time T . We assume that the customer arrived in the system at time $t < T$ creates a point in the screened process with probability

$$S(t) = 1 - B(T - t)$$

or does not create it with probability $1 - S(t)$. We name the points occurred in the screened process before t as customers in the screened process at time t .

Let us denote the customers number in the screened process at the moment t by $n(t)$. Then, if at the initial moment $t_0 < T$ the system is empty, we have the following equality at the moment T :

$$P\{i(T) = m\} = P\{n(T) = m\}.$$

Note that this method exactly determines the characteristics of the process $V(t)$ since the screened process contains only customers which do not finish the service at the moment T .

3. The system of Kolmogorov differential equations

Let us consider the three-dimensional Markovian process $\{k(t), n(t), V(t)\}$, where $k(t)$ identifies the state of the modulating Markov chain of the MMPP input process at time t ($1 \leq k(t) \leq K$), which is defined through the infinitesimal generator matrix \mathbf{Q} and rate matrix $\mathbf{\Lambda}$:

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1K} \\ q_{21} & q_{22} & \dots & q_{2K} \\ \dots & \dots & \dots & \dots \\ q_{K1} & q_{K2} & \dots & q_{KK} \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_K \end{bmatrix}.$$

Denoting the probability distribution of this process by $P(k, n, z, t) = P\{k(t) = k, n(t) = n, V(t) < z\}$, we can write the corresponding system of Kolmogorov differential equations taking into account the formula of total probability:

$$\begin{aligned} P(k, n, z, t + \Delta t) &= P(k, n, z, t) (1 - \lambda_k) (1 - q_{kk}) + \\ &P(k, n, z, t) \lambda_k \Delta t (1 - S(t)) + \\ &\lambda_k \Delta t S(t) \int_0^z P(k, n - 1, z - y, t) dG(y) + \\ &\sum_{\nu \neq k} q_{\nu k} \Delta t P(\nu, n, z, t) + o(\Delta t), \end{aligned} \tag{1}$$

for $k = 1 \dots K, n = 0, 1, 2, \dots, z > 0$.

From (1), we obtain the system of Kolmogorov differential equations

$$\begin{aligned} \frac{\partial P(k, n, z, t)}{\partial t} &= \lambda_k S(t) \left[\int_0^z P(k, n - 1, z - y, t) dG(y) \right] + \\ &\sum_{\nu} q_{\nu k} P(\nu, n, z, t) \end{aligned}$$

for $k = 1 \dots K, n = 0, 1, 2, \dots, z > 0$.

We introduce a partial characteristic function of the form:

$$\begin{aligned} H(k, u_1, u_2, t) &= M \{ \exp(ju_1 n(t) + ju_2 V(t)) \} = \\ &\sum_{n=0}^{\infty} e^{ju_1 n} \int_0^{\infty} e^{ju_2 z} P(k, n, z, t) dz \end{aligned}$$

for $k = 1 \dots K$, $n = 0, 1, 2, \dots$, $z > 0$.

Considering that

$$\sum_{n=0}^{\infty} e^{ju_1 n} \int_0^{\infty} e^{ju_2 z} \int_0^z P(k, n-1, z-y, t) dG(y) dz = e^{ju_1} H(k, u_1, u_2, t) G^*(u_2),$$

where

$$G^*(u_2) = \int_0^{\infty} e^{ju_2 y} dG(y), \quad (2)$$

we can write the following system of equations:

$$\frac{\partial H(k, u_1, u_2, t)}{\partial t} = \lambda_k S(t) H(k, u_1, u_2, t) [e^{ju_1} G^*(u_2) - 1]$$

for $k = 1 \dots K$.

We write this system in the form of a matrix equation

$$\frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) [\mathbf{\Lambda} S(t) (e^{ju_1} G^*(u_2) - 1) + \mathbf{Q}] \quad (3)$$

with the initial condition

$$\mathbf{H}(u_1, u_2, t_0) = \mathbf{r}, \quad (4)$$

where

$$\mathbf{H}(u_1, u_2, t) = [H(1, u_1, u_2, t), H(2, u_1, u_2, t), \dots, H(K, u_1, u_2, t)],$$

and

$$\mathbf{r} = [r(1), r(2), \dots, r(K)],$$

is the row vector of the stationary distribution of the modulating Markov chain:

$$\begin{cases} \mathbf{rQ} = 0, \\ \mathbf{re} = 1, \end{cases} \quad (5)$$

\mathbf{e} being a column vector with all entries equal to 1.

4. The asymptotic analysis method

The exact solution of the equation (3) is, in general, not available, but it is possible to get asymptotic results in case of heavy loads. To this aim we will use the asymptotic analysis method under the condition of an infinitely growing arrival rate. Let us substitute $\mathbf{\Lambda} = N\mathbf{\Lambda}^1$ and $\mathbf{Q} = N\mathbf{Q}^1$ into the equation (3), where N is some parameter which will be used for the asymptotic analysis ($N \rightarrow \infty$ in theoretical studies).

Then, the equation (3) takes the form

$$\frac{1}{N} \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) [\mathbf{\Lambda}^1 S(t) (e^{ju_1} G^*(u_2) - 1) + \mathbf{Q}^1] \quad (6)$$

with the initial condition

$$\mathbf{H}(u_1, u_2, t_0) = \mathbf{r}. \quad (7)$$

4.1. The first-order asymptotic analysis

The main result is summarized by the following lemma.

Lemma. *The first-order asymptotic characteristic function of the probability distribution of the process $\{k(t), n(t), V(t)\}$ has the form*

$$\mathbf{H}(u_1, u_2, t) = \mathbf{r} \exp \left\{ N\lambda [ju_1 + ju_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\},$$

where the row vector \mathbf{r} is defined by the system of linear equations (5), λ denotes the average rate

$$\lambda = \mathbf{r}\mathbf{\Lambda}\mathbf{e}$$

and a_1 is the mean of the random variable defining the customer capacity

$$a_1 = \int_0^{\infty} y dG(y).$$

Proof.

Let us perform the substitutions

$$\varepsilon = \frac{1}{N}, u_1 = \varepsilon w_1, u_2 = \varepsilon w_2, \mathbf{H}(u_1, u_2, t) = \mathbf{F}_1(w_1, w_2, t, \varepsilon) \quad (8)$$

in the expressions (5) and (6).

Then the problem (5) – (6) takes the form

$$\varepsilon \frac{\partial \mathbf{F}_1(w_1, w_2, t, \varepsilon)}{\partial t} =$$

$$\mathbf{F}_1(w_1, w_2, t, \varepsilon) [\mathbf{\Lambda}^1 S(t) (e^{j\varepsilon w_1} G^*(\varepsilon w_2) - 1) + \mathbf{Q}^1] \quad (9)$$

with the initial condition

$$\mathbf{F}_1(w_1, w_2, t, \varepsilon) = \mathbf{r}.$$

Let us find the asymptotic solution (where $\varepsilon \rightarrow 0$) of the problem (8) – (9), i.e. the $\mathbf{F}_1(w_1, w_2, t) = \lim_{\varepsilon \rightarrow 0} \mathbf{F}_1(w_1, w_2, t, \varepsilon)$.

Step 1. Letting $\varepsilon \rightarrow 0$ in (9), we obtain

$$\mathbf{F}_1(w_1, w_2, t) \mathbf{Q}^1 = 0.$$

Comparing this equation with the first one in (5), we can conclude that $\mathbf{F}_1(w_1, w_2, t)$ can be expressed as

$$\mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \Phi_1(w_1, w_2, t), \quad (10)$$

where $\Phi_1(w_1, w_2, t)$ is some scalar function which satisfies the condition

$$\Phi_1(w_1, w_2, t_0) = 1.$$

Step 2. Let us multiply (9) by vector \mathbf{e} , substitute (10), divide the results by ε and perform the asymptotic transition $\varepsilon \rightarrow 0$. Then, taking into account that $\mathbf{Q}^1 \mathbf{e} = 0$ and $\mathbf{r} \mathbf{e} = 1$, we obtain the following differential equation for the function $\Phi_1(w_1, w_2, t)$

$$\frac{\partial \Phi_1(w_1, w_2, t)}{\partial t} = \Phi_1(w_1, w_2, t) [\lambda S(t) (jw_1 + jw_2 a_1)]. \quad (11)$$

The solution of (11) with the initial condition gives

$$\Phi_1(w_1, w_2, t) = \exp \left\{ \lambda (jw_1 + jw_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}.$$

Substituting this expression into (10), we obtain

$$\mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \exp \left\{ \lambda (jw_1 + jw_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}.$$

Using substitutions (8), we can write the asymptotic (as $\varepsilon \rightarrow 0$) equality:

$$\begin{aligned} \mathbf{H}(u_1, u_2, t) &= \mathbf{F}_1(w_1, w_2, t, \varepsilon) \approx \mathbf{F}_1(w_1, w_2, t) = \mathbf{r}\Phi_1(w_1, w_2, t) = \\ &= \mathbf{r} \exp \left\{ \lambda \left[j \frac{u_1}{\varepsilon} + j \frac{u_2}{\varepsilon} a_1 \right] \int_{t_0}^t S(\tau) d\tau \right\} = \\ &= \mathbf{r} \exp \left\{ N\lambda [ju_1 + ju_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\}. \end{aligned}$$

The proof is complete.

Corollary. When $t = T$ we obtain the characteristic function of the process $\{i(t), V(t)\}$ in the steady state regime

$$H(u_1, u_2, t) = \exp \{ N\lambda b_1 [ju_1 + ju_2 a_1] \},$$

where

$$b_1 = \int_{-\infty}^T S(\tau) d\tau = \int_{-\infty}^T (1 - B(T - \tau)) d\tau = \int_0^{\infty} (1 - B(\tau)) d\tau$$

denotes the mean service time.

4.2. The second-order asymptotic analysis

The main result is summarized by the following theorem.

Theorem. The second-order asymptotic characteristic function of the probability distribution of the process $\{k(t), n(t), V(t)\}$ has the form

$$\begin{aligned} \mathbf{H}(u_1, u_2, t) &= \mathbf{r} \exp \left\{ N\lambda (ju_1 + ju_2 a_1) \int_{t_0}^t S(\tau) d\tau + \right. \\ &\quad \left. \frac{(ju_1)^2}{2} \left(N\lambda \int_{t_0}^t S(\tau) d\tau + N\kappa \int_{t_0}^t S^2(\tau) d\tau \right) + \right. \\ &\quad \left. \frac{(ju_2)^2}{2} \left(N\lambda a_2 \int_{t_0}^t S(\tau) d\tau + N\kappa a_1^2 \int_{t_0}^t S^2(\tau) d\tau \right) + \right. \end{aligned}$$

$$j^2 u_1 u_2 \left(N \lambda a_1 \int_{t_0}^t S(\tau) d\tau + N \kappa a_1 \int_{t_0}^t S^2(\tau) d\tau \right) \Bigg\},$$

where

$$\kappa = 2\mathbf{g}(\mathbf{\Lambda}^1 - \lambda\mathbf{I})\mathbf{e},$$

and the row vector \mathbf{g} satisfies the linear matrix system

$$\mathbf{g}\mathbf{Q}^1 = \mathbf{r}(\lambda\mathbf{I} - \mathbf{\Lambda}^1),$$

$$\mathbf{g}\mathbf{e} = 1.$$

Proof.

Denote by $\mathbf{H}_2(u_1, u_2, t)$ a multi-dimensional function that satisfies the equation

$$\mathbf{H}(u_1, u_2, t) = \mathbf{H}_2(u_1, u_2, t) \exp \left\{ N \lambda (j u_1 + j u_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}. \quad (12)$$

Substituting this expression into (6) and (7), we obtain the following problem:

$$\begin{aligned} \frac{1}{N} \frac{\partial \mathbf{H}_2(u_1, u_2, t)}{\partial t} + \lambda (j u_1 + j u_2 a_1) S(t) \mathbf{H}_2(u_1, u_2, t) &= \quad (13) \\ &= \mathbf{H}_2(u_1, u_2, t) [\mathbf{\Lambda}^1 S(t) (e^{j u_1} G^*(u_2) - 1) + \mathbf{Q}^1], \end{aligned}$$

with the initial condition

$$\mathbf{H}_2(u_1, u_2, t_0) = \mathbf{r}. \quad (14)$$

Let us perform the following substitutions

$$\varepsilon^2 = \frac{1}{N}, u_1 = \varepsilon w_1, u_2 = \varepsilon w_2, \mathbf{H}_2(u_1, u_2, t) = \mathbf{F}_2(w_1, w_2, t, \varepsilon). \quad (15)$$

Using these notations the problem (13) – (14) can be rewritten in the form

$$\begin{aligned} \varepsilon^2 \frac{\partial \mathbf{F}_2(w_1, w_2, t, \varepsilon)}{\partial t} + \mathbf{F}_2(w_1, w_2, t, \varepsilon) \lambda (j \varepsilon w_1 + j \varepsilon w_2 a_1) S(t) &= \quad (16) \\ &= \mathbf{F}_2(w_1, w_2, t, \varepsilon) [\mathbf{\Lambda}^1 S(t) (e^{j \varepsilon w_1} G^*(\varepsilon w_2) - 1) + \mathbf{Q}^1], \end{aligned}$$

with the initial condition

$$\mathbf{F}_2(w_1, w_2, t_0, \varepsilon) = \mathbf{r}. \quad (17)$$

Let us find the asymptotic solution (as $\varepsilon \rightarrow 0$) of this problem, i.e. the function $\mathbf{F}_2(w_1, w_2, t) = \lim_{\varepsilon \rightarrow 0} \mathbf{F}_2(w_1, w_2, t, \varepsilon)$.

Step 1. Letting $\varepsilon \rightarrow 0$ in (16) – (17), we obtain the following system of equations:

$$\begin{cases} \mathbf{F}_2(w_1, w_2, t) \mathbf{Q}^1 = \mathbf{0}, \\ \mathbf{F}_2(w_1, w_2, t_0) = \mathbf{r}, \end{cases}$$

Then, using (4), we can write

$$\mathbf{F}_2(w_1, w_2, t) = \mathbf{r}\Phi_2(w_1, w_2, t), \tag{18}$$

where $\Phi_2(w_1, w_2, t)$ is some scalar function which satisfies the condition

$$\Phi_2(w_1, w_2, t_0) = 1.$$

Step 2. Using (18), the function $\mathbf{F}_2(w_1, w_2, t)$ can be represented in the expansion form

$$\mathbf{F}_2(w_1, w_2, t, \varepsilon) = \Phi_2(w_1, w_2, t) [\mathbf{r} + \mathbf{g}(j\varepsilon w_1 + j\varepsilon w_2 a_1) S(t)] + \mathbf{O}(\varepsilon^2), \tag{19}$$

where \mathbf{g} is some row vector which satisfying the condition $\mathbf{g}\mathbf{e} = 1$ and $\mathbf{O}(\varepsilon^2)$ is row vector whose elements are infinitesimals of the same order as ε^2 .

Let us use the substitution (19) and the Taylor-Maclaurin expansions

$$e^{j\varepsilon w_1} = 1 + j\varepsilon w_1 + O(\varepsilon^2), e^{j\varepsilon w_2} = 1 + j\varepsilon w_2 + O(\varepsilon^2)$$

in (16). Considering the (2), we perform in the obtained equality of the limiting transition $\varepsilon \rightarrow 0$, we obtain matrix equation for the vector \mathbf{g}

$$\mathbf{g}\mathbf{Q}^1 = \mathbf{r}(\lambda\mathbf{I} - \mathbf{\Lambda}^1),$$

where \mathbf{I} is diagonal unit matrix.

Step 3. We multiply the (16) by the \mathbf{e} , using (19) and the second-order expansions

$$e^{j\varepsilon w_1} = 1 + j\varepsilon w_1 + \frac{(j\varepsilon w_1)^2}{2} + O(\varepsilon^2),$$

$$e^{j\varepsilon w_2} = 1 + j\varepsilon w_2 + \frac{(j\varepsilon w_2)^2}{2} + O(\varepsilon^2).$$

As a result of simple transformations with the notation

$$\kappa = 2\mathbf{g}(\mathbf{\Lambda}^1 - \lambda\mathbf{I})\mathbf{e},$$

we obtain the following differential equation for the function $\Phi_2(w_1, w_2, t)$

$$\frac{\partial \Phi_2(w_1, w_2, t)}{\partial t} = \Phi_2(w_1, w_2, t) \left\{ \frac{(jw_1)^2}{2} (\lambda S(t) + \kappa S^2(t)) + \frac{(jw_2)^2}{2} (\lambda a_2 S(t) + \kappa a_1^2 S^2(t)) + j^2 w_1 w_2 (\lambda a_1 S(t) + \kappa a_1 S^2(t)) \right\},$$

where $a_2 = \int_0^\infty y^2 dG(y)$ is the second moment of the random customer capacity ν .

The solution of the latter equation with the available initial condition $\Phi_2(w_1, w_2, t_0) = 1$ gives the expression $\Phi_2(w_1, w_2, t)$

$$\Phi_2(w_1, w_2, t) = \exp \left\{ \frac{(jw_1)^2}{2} \left(\lambda \int_{t_0}^t S(\tau) d\tau + \kappa \int_{t_0}^t S^2(\tau) d\tau \right) + \frac{(jw_2)^2}{2} \left(\lambda a_2 \int_{t_0}^t S(\tau) d\tau + \kappa a_1^2 \int_{t_0}^t S^2(\tau) d\tau \right) + j^2 w_1 w_2 \left(\lambda a_1 \int_{t_0}^t S(\tau) d\tau + \kappa a_1 \int_{t_0}^t S^2(\tau) d\tau \right) \right\},$$

and substituting in (18) we obtain

$$\mathbf{F}_2(w_1, w_2, t) = \mathbf{r} \exp \left\{ \frac{(jw_1)^2}{2} \left(\lambda \int_{t_0}^t S(\tau) d\tau + \kappa \int_{t_0}^t S^2(\tau) d\tau \right) + \frac{(jw_2)^2}{2} \left(\lambda a_2 \int_{t_0}^t S(\tau) d\tau + \kappa a_1^2 \int_{t_0}^t S^2(\tau) d\tau \right) + j^2 w_1 w_2 \left(\lambda a_1 \int_{t_0}^t S(\tau) d\tau + \kappa a_1 \int_{t_0}^t S^2(\tau) d\tau \right) \right\}. \quad (20)$$

Performing in (20) the substitutions inverse to (15) and (12), we obtain the following expression for the asymptotic characteristic function of the

number of customers of screened process and total capacity of customers at the moment t :

$$\mathbf{H}(u_1, u_2, t) = \mathbf{r} \exp \left\{ N\lambda (ju_1 + ju_2a_1) \int_{t_0}^t S(\tau) d\tau + \frac{(ju_1)^2}{2} \left(N\lambda \int_{t_0}^t S(\tau) d\tau + N\kappa \int_{t_0}^t S^2(\tau) d\tau \right) + \frac{(ju_2)^2}{2} \left(N\lambda a_2 \int_{t_0}^t S(\tau) d\tau + N\kappa a_1^2 \int_{t_0}^t S^2(\tau) d\tau \right) + j^2 u_1 u_2 \left(N\lambda a_1 \int_{t_0}^t S(\tau) d\tau + N\kappa a_1 \int_{t_0}^t S^2(\tau) d\tau \right) \right\},$$

where

$$\kappa = 2\mathbf{g}(\mathbf{\Lambda}^1 - \lambda\mathbf{I})\mathbf{e},$$

and the row vector \mathbf{g} satisfies the linear matrix system

$$\mathbf{g}\mathbf{Q}^1 = \mathbf{r}(\lambda\mathbf{I} - \mathbf{\Lambda}^1),$$

$$\mathbf{g}\mathbf{e} = 1.$$

The proof is complete.

Corollary 1. *When $t = T$ we obtain the characteristic function of the process $i(t), V(t)$ in the steady state regime*

$$H(u_1, u_2, t) = \exp \left\{ N\lambda (ju_1 + ju_2a_1) b_1 + \frac{(ju_1)^2}{2} (N\lambda b_1 + N\kappa b_2) + \frac{(ju_2)^2}{2} (N\lambda a_2 b_1 + N\kappa a_1^2 b_2) + j^2 u_1 u_2 (N\lambda a_1 b_1 + N\kappa a_1 b_2) \right\}, \quad (21)$$

where

$$b_2 = \int_{-\infty}^T S^2(\tau) d\tau.$$

From the form of function (21) it is clear that the two-dimensional process $i(t), V(t)$ is asymptotically Gaussian with the vector of mathematical expectations

$$\mathbf{a} = [N\lambda b_1, N\lambda a_1 b_1]$$

and the covariance matrix

$$\mathbf{A} = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} N\lambda b_1 + N\kappa b_2 & N\lambda a_1 b_1 + N\kappa a_1 b_2 \\ N\lambda a_1 b_1 + N\kappa a_1 b_2 & N\lambda a_2 b_1 + N\kappa a_1^2 b_2 \end{bmatrix}.$$

Corollary 2. *The asymptotic characteristic function of the total customer capacity in the steady-state regime is given by a Gaussian characteristic function*

$$H(u, t) = \exp \left\{ juN\lambda a_1 b_1 + \frac{(ju)^2}{2} (N\lambda a_2 b_1 + Na_1^2 \kappa b_2) \right\},$$

with parameters $a = N\lambda a_1 b_1$ and $\sigma^2 = N\lambda a_2 b_1 + Na_1^2 \kappa b_2$.

5. Conclusions

In the paper, a queueing system with random customers capacity and service time independent of its capacity is considered in case of correlated arrivals, described by an MMPP process. For such system, the total customers capacity is derived by using the asymptotic analysis method in case of heavy loads. It is obtained that the stationary probability distribution of total capacity can be approximated by a Gaussian distribution and the parameters of the approximation are derived in the paper.

References

1. *Tikhonenko, O.M.* Destricted capacity queueing systems. Determination of their characteristics *Autom. Remote Control*, 58 (6), 969-972 (1997)
2. *Tikhonenko, O.M.* Computer Systems Probability Analysis. Akademicka Oficyna Wydawnicza EXIT, Warsaw (2006) (in Polish)
3. *Romm, E.L., Skitovich, V.V.* On certain generalization of problem of Erlang. *Avtomat. i Telemekh.*, No. 6, 164-168 (1971)
4. *Tikhonenko, O.M.* Systems for servicing requests of random length with restrictions. *Avtomat. i Telemekh.*, No. 10, 126-134 (1991)
5. *Tikhonenko, O.M.* Queueing Systems with Processor Sharing and Limited Resources. *Autom. Remote Control*, No. 5, 803-815 (2010)
6. *Heyman, D.P., Lucantoni, D.* Modelling multiple IP traffic streams with rate limits. *IEEE/ACM Trans. on Networking*, No. 11, 948-958 (2003)

7. *Chakravarthy, S.R.* Markovian arrival processes. Wiley Encyclopedia of Operations Research and Management Science (2010)
8. *Moiseev, A., Nazarov, A.* Asymptotic Analysis of the Infinite-Server Queueing System with High-Rate Semi-Markov Arrivals. IEEE International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT 2014), IEEE Press, New York, 507-513 (2014)
9. *Naumov, V., Samuoylov, K., Sopin, E., Yarkina, N., Andreev, S., Samuylov, A.* LTE Performance Analysis Using Queueing Systems with Finite Resources and Random Requirements. Proc. of 8th Int. Congress on Ultra Modern Telecommunications and Control Systems and Workshops, 100-103 (2015).
10. *Moiseev A.N., Nazarov A.A.* Infinite-Server Queueing System and Networks. Tomsk. Publishing house STL (2015) (in Russian)

УДК 004.932

Исследование взаимосвязи субъективных оценок качества восприятия видео и параметра Хёрста

М. А. Маколкина, Е. В. Суродеева

*Санкт-Петербургский государственный университет Телекоммуникаций
им. М.А. Бонч-Бруевича*

Аннотация. Сегодня в сети появляется большое количество услуг и приложений, ориентированных на передачу видеотрафика, в тоже время присутствует трафик и от других приложений. Необходимость передавать по сетям трафик разной природы требует пересмотра подходов к проектированию и управлению сетями связи. Также стоит выделить оценку качества восприятия предоставляемых услуг, поскольку сегодня пользователи уделяют этому вопросу пристальное внимание. Это приводит к сложной задаче формирования новых показателей Quality of Experience (QoE) и разработке методов оценки QoE услуг. В статье исследуется метод оценки качества передачи видео на основе анализа значений параметра Хёрста и их взаимосвязи с субъективными оценками качества передачи видео.

Ключевые слова: видеотрафик, IPTV услуги, качество обслуживания, качество восприятия, субъективная оценка качества, параметр Хёрста, самоподобие.

1. Введение

В современном мире телекоммуникаций каждый день появляется новое приложение, включающее в себя не просто трафик разного типа, но и отличающееся разной природой и характеристиками трафика [1, 2, 3]. Так, большинство приложений сегодня ориентированы на передачу видео, в зависимости от оконечных терминалов пользователей и типа услуг, одно и тоже видео формируется в различные транспортные потоки с отличающимися характеристиками передачи. Также к видео могут добавляться дополнительные данные, например, если речь идет об услугах «дополненной реальности» [4, 5, 6] возможен трафик от сенсорных узлов, аудио, текст или трехмерное изображение. Предоставляемые операторами связи в настоящее время IPTV услуги широко внедрены и пользуются популярностью, но для того, чтобы привлечь большее количество пользователей операторы постоянно модернизируют существующие и внедряют новые услуги [7, 8]. Так, с появлением на рынке камер «дополненной реальности» появляются новые возможности по мониторингу предпочтений пользователей, управлением видеопотоками, упрощению навигации по интерфейсу и оценке качества восприятия услуги. Более того, речь идет о возможности оценить качество восприятия, не только анализируя объективные параметры, но и захватить субъективные параметры, которые до недавнего времени вообще не брались в расчет. Так эмоции, настроение, культурный

фон, степень удовлетворенности пользователя услугой, информативность контента сегодня реально оценить. Следующий вопрос, который неизменно встает, а по каким критериям нам оценивать эмоции и информативность контента, и имеют ли они взаимосвязь с объективными параметрами, которые каждый оператор хорошо научился считать, анализировать и реагировать на их изменения. В статье рассматривается подход к оценке качества восприятия, основанный на установленной взаимосвязи субъективных оценок качества восприятия и параметра Хёрста [9, 10]. В ходе исследования рассмотрены видеопотоки с различными характеристиками видеокодеков, подвергающихся ухудшающему воздействию со стороны сети и при добавлении других типов трафика и произведена субъективная оценка качества передачи видео и расчет параметра Хёрста.

2. Цель исследования

На сегодняшний день для оценки качества предоставления видеослуж не разработана методика, которая бы давала однозначный ответ, удовлетворен пользователь услугой или нет. По аналогии с E-моделью, разработанной для оценки качества передачи речи, предложен метод оценки качества передачи видео на основе анализа степени самоподобия трафика. Результатом расчета E-модели является значение R-фактора, включающего в себя порядка 20 оцениваемых параметров. Значения R-фактора имеют однозначное соответствие с субъективными оценками пользователей. Целью данного исследования является установление зависимостей между субъективными оценками качества передачи видео и значениями параметра Хёрста для видео, закодированного в разных форматах, и подвергающегося ухудшающему воздействию сети.

3. Модельная сеть

Для проведения исследований была собрана модельная сеть, представленная на рисунке 1. Модельная сеть состоит из комплекса IPTV, который принимает со спутника сигнал. С приемника сигнал попадает на транскодер, где происходит его преобразование в различные форматы. С транскодера сигнал поступает на сервер Middleware, где подвергается обработке для последующего воспроизведения пользователю, например, для телевизионных каналов добавляется электронная программа телепередач (Electronic program guide, EPG), канал закрепляется за определенным пакетом каналов, назначается возрастной рейтинг, производится описание канала и назначение стоимости просмотра и т.д. Далее сигнал с мультиплектора поступает на коммутатор, откуда в режиме multicast поступает на set-top-box пользователя.

При прохождении сигнала через сегмент сети на него осуществляется воздействие с помощью сервера NetDisturb. NetDisturb – это

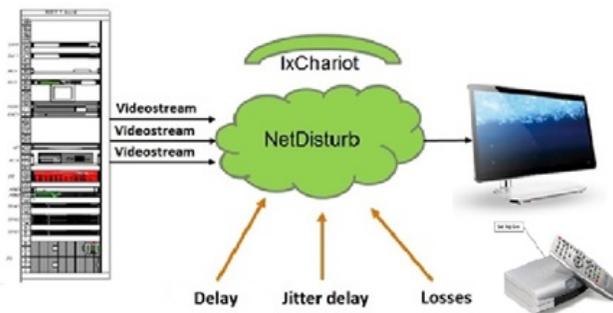


Рис. 1. Структура модельной сети

открытый программный продукт, который позволяет произвести ухудшения на IP-сети, тем самым влияя на её сетевые характеристики и работоспособность. В программе можно задать тип потерь и величину, рядом задаются тип задержки и ее величина в одну сторону, и в обратную сторону. Сервер с установленной программой подключается в сеть между источником и получателем и работает в режиме прозрачного канала. Видеопоток при прохождении через подвергается влиянию NetDisturb, который искусственно вносит задержки и потери, далее на стороне пользователя оценивается параметр Хёрста с помощью метода графика анализа дисперсии и выставляется оценка качества восприятия группой экспертов [11]. В качестве метода субъективной оценки передачи видео был выбран метод SSCQE (Single Stimulus Continuous Quality Evaluation, Метод непрерывной оценки качества при одном источнике воздействия), описанный в рекомендации МСЭ ВТ.500-13 [12]. Этот метод используется для оценки качества передачи видеоизображений в условиях близким к домашним условиям просмотра видео пользователями, когда нет возможности сравнить тестируемое видео с эталонным, полученным напрямую с источника, т.е. от телерадиокомпании, формирующей контент. Для оценки качества используется непрерывная шкала. Наблюдателям предлагают оценить общее качество видео изображения, проставив отметку на вертикальной шкале. Шкала представляет собой систему непрерывной оценки во избежание ошибок квантования, но она разделена на пять отрезков равной длины, соответствующих обычной пятибалльной шкале качества МСЭ-R. Термины, определяющие различные уровни являются теми же, что используются обычно: отлично, хорошо, удовлетворительно,

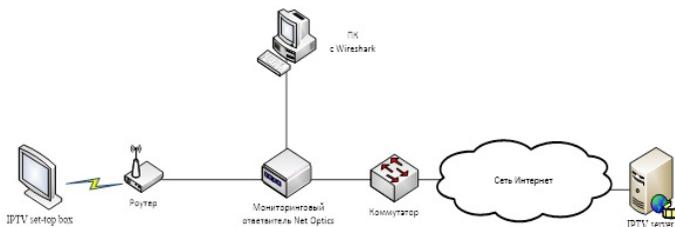


Рис. 2. Схема стенда для исследования трафика IPTV

плохо, неприемлемо. В качестве второго субъективного метода оценки был выбран метод ACR (Absolute Category Rating), подробно описанный в Рекомендации МСЭ-Т Р.910 [13]. Данный метод использует категориальные оценки. Тестовая последовательность представляется один раз и оценивается по установленной шкале оценок. Преимуществом метода ACR является способность оценивать только принятую последовательность у получателя, не имея эталонной, что позволяет приблизиться к реальным условиям работы сети и оценкам конечных пользователей.

4. Результаты исследования

В результате эксперимента был получен IPTV трафик в исходящем и входящем направлениях. Также был проведен ряд экспериментов с добавлением в сеть трафика от других источников, для того, чтобы оценить взаимное влияние трафика при ограниченности сетевого ресурса. Схема стенда представлена на рис. 2.

Эксперимент состоял из следующих этапов:

- Переключить устройство перехвата трафика в активный режим;
- Послать на приставку IPTV видеопоток;
- С помощью устройства перехвата трафика, происходит захват пакетов генерируемых сервером IPTV.

Коэффициент Хёрста рассчитан методом анализа изменения дисперсии для трафика при просмотре выпуска новостей на «Первом канале». Оценка коэффициента Хёрста для исходящего и входящего трафиков IPTV иллюстрируется на рисунках 3 и 4. Наклон линии регрессии указывает на то, что для входящего трафика коэффициент Хёрста составляет 0.57, а для исходящего трафика – 0.52. Таким образом, трафик является самоподобным с малой скоростью самоподобия.

Результаты расчетов сведены в таблицу 1.

Таким образом, полученные результаты показали, что трафики BitTorrent и M2M в исходящем и входящем направлениях обладают

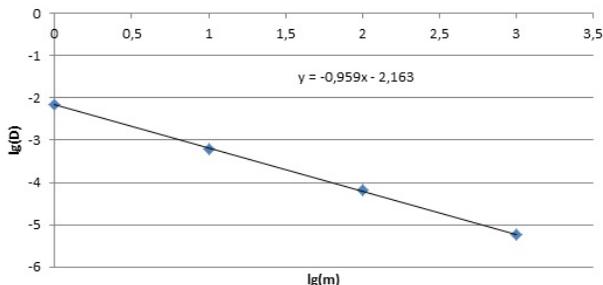


Рис. 3. Оценка коэффициента Хёрста методом анализа изменений дисперсии для исходящего трафика IPTV

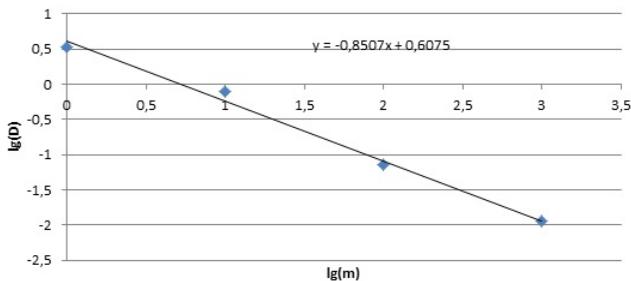


Рис. 4. Оценка коэффициента Хёрста методом анализа изменений дисперсии для входящего трафика IPTV

особенностями, характерными для антиперсистентного потока, которые могут в значительной степени влиять на качество предоставляемых услуг. Субъективная оценка. В ходе проведения эксперимента анализировалось видео в разных форматах от разных источников при внесении изменений в работу сети. В эксперименте участвовало 15 человек, длительность просмотра видео составляла 1,5 минуты, оценки выставлялись по пятибалльной шкале, параллельно рассчитывался параметр Хёрста. Сводные результаты представлены в таблице 2.

Из полученных результатов видно, что параметр Хёрста взаимосвязан с субъективными оценками качества передачи видео и может использоваться для анализа степени удовлетворенности пользователя видеослугами.

Таблица 1

Значения коэффициента Хёрста

Тип трафика	Направление	Коэфф. Херста	Свойства
Skype	Входящий	0,545	Самоподобный
Skype	Исходящий	0,54	Самоподобный
Torrent	Входящий	0,36	Антиперсиситентный
Torrent	Исходящий	0,38	Антиперсиситентный
M2M	Входящий	0,38	Антиперсиситентный
M2M	Исходящий	0,21	Антиперсиситентный
IPTV	Входящий	0,57	Самоподобный
IPTV	Исходящий	0,52	Самоподобный

Таблица 2

Оценка параметра Хёрста (H) и субъективная оценка для разных форматов видео (MOS)

Хорошее качество		Среднее качество		Низкое качество	
MOS	H	MOS	H	MOS	H
1,55	0,733	1,92	0,642	1,92	0,635
2,73	0,758	2,83	0,635	2,51	0,715
3,26	0,796	4,08	0,677	3,74	0,767
3,55	0,812	3,96	0,689	3,52	0,741
3,98	0,895	4,41	0,692	4,16	0,804

5. Выводы

Предложена модельная сеть, на базе которой проведена оценка качества передачи видео на примере трафика IPTV с добавлением в сеть других типов трафика.

В дальнейшем планируется расширить модельную сеть, добавив другие виды видеопотоков, характеристики и типы M2M трафика, исследовать их взаимное влияние, влияние на сеть и оценить качество восприятия передаваемого трафика, установить характер взаимосвязи между субъективными оценками QoE, характеристики сети, свойствами потока и эмоциями пользователей.

Предлагается использование параметра Хёрста в качестве объективной метрики качества восприятия видео по аналогии с R-фактором

для качества передачи речи. На основе взаимосвязи параметра Хёрста и субъективных методов оценки качества восприятия видео может быть разработан метод объективной оценки качества передачи видео.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00209 мол а "Разработка принципов интеграции технологии дополненной реальности и Интернета Вещей."

Литература

1. *Paramonov, A., Koucheryavy A.* M2M traffic models and flow types in case of mass event detection (2014) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8638 LNCS, pp. 294-300.
2. State of the art and research challenges for public flying ubiquitous sensor networks. Koucheryavy A., Vladyko A., Kirichek R. Lecture Notes in Computer Science. 2015. Т. 9247. С. 299-308.
3. The home network traffic models investigation. Kirichek R., Golubeva M., Kulik V., Koucheryavy A. В сборнике: 18th International Conference on Advanced Communication Technology (ICACT) 2016. С. 97-100.
4. *Bergenti F., Danilo Gotta.* «Augmented Reality for Field Maintenance of Large Telecommunication Networks», Conference and Exhibition of the European Association of Virtual and Augmented Reality, 2014/
5. *Ganapati, Priya.* "How it Works: Augmented Reality."Wired. Aug. 25, 2009. <http://www.wired.com/gadgetlab/2009/08/total-immersion/>
6. *Jonathan J. Hull, Berna Erol, Jamey Graham, Qifa Ke, Hidenobu Kishi, Jorge Moraleda, Daniel G. Van Olst.* "Paper-Based Augmented Reality". 17th International Conference on Artificial Reality and Telexistence, 2007.
7. *Chang S., Vetro A.* Video adaptation: concepts, technologies, and open issues / to appear in Proceedings of the IEEE. — January 2005.
8. *Karl-Heinz Wenisch, Rainer Link.* "IPTV as an Alternative to Coax. Advantages of an IPTV-system for enterprises". www.teracue.com. September, 2014.
9. *Makolkina, M., Prokopiev, A., Paramonov, A., Koucheryavy, A.* The quality of experience subjective estimations and the Hurst parameters values interdependence (2014) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8638 LNCS, pp. 311-318.
10. Estimation quality parameters of transferring image and voice data over Zigbee in transparent mode. Kirichek R., Makolkina M., Sene J., Takhtuev V. Communications in Computer and Information Science. 2016. Т. 601. С. 260-267.

11. *M.Z.Shafiq and all.* A First Look at Cellular Machine-to-Machine Traffic: Large Scale Measurement and Characterization. 12th ACM Sigmetrics Performance International Conference. June 11-15, London, England, UK, 2012.
12. ITU-R Recommendation BT.500-13. Methodology for the Subjective Assessment of the Quality of Television Pictures. Geneva, January, 2012.
13. ITU-T Recommendation P.913. Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment. Geneva, January, 2014.

UDC 004.932

Study of the interrelationship of subjective perception of the video quality and Hurst parameter of traffic

M. A. Makolkina, E. V. Surodeeva

Department of Infocommunication Systems , The Bonch-Bruевич Saint-Petersburg State University of Telecommunications, str. pr. Bolsheviki, d.22, St. Petersburg, Russia, 193232

The paper is dedicated to study of quality of video services. The goal of the study to describe quality of service by any traffic characteristic. Results of investigation shows relation between the subjective quality estimations and the Hurst parameter of the video traffic.

Keywords: video traffic, IPTV, quality of service, subjective quality estimation, Hurst parameter, selfsimilarity.

Speed-up estimation of a system with random volume customers

E. V. Morozov^{*†}, L. V. Potakhina^{*†}

** Institute of Applied Mathematical Research, Karelian Research Centre
Pushkinskaya str. 11, Petrozavodsk, 185035, Russia*

*† Petrozavodsk State University
Lenina str. 33, Petrozavodsk, 185910, Russia*

Abstract. We consider a general single-server system in which each customer has service time and also a random volume, and the total volume (system capacity) is upper bounded by a finite constant. To estimate the moments of the lost customer volume, we apply an approximation using the renewal theory. To illustrate the accuracy of the approximation we develop regenerative simulation. Moreover, we use RESTART/splitting technique to speed-up estimation based on simulation in the case when system capacity is large and loss becomes a rare event.

Keywords: Queueing system, random volume customer, regeneration, simulation, splitting.

1. Introduction

A finite volume of the computer memory is very important aspect of the modern high performance systems. The analysis of finite space systems, in which customers have both random service time and random volume is quite important to study various problems related to the functioning of high performance computer and communication systems [1, 2]. The total accumulated volume can be considered as the basic process describing dynamics of the system. As a basic QoS parameter we consider the volume of the lost customer and estimate its moments. An important role in this analysis plays an analogy between the process of the lost customers and an associated renewal process generated by the customer volumes. This analogy is based on the so-called inspection paradox and implies a useful approximation of the mean lost volume. To study the accuracy of the approximation, we perform regenerative simulation. Moreover, when the volume capacity is large enough, a loss of customer becomes a rare event, and crude Monte Carlo simulation becomes ineffective requiring a huge simulation time to accumulate the number of observations which is sufficient to estimate required probability with a given accuracy. To overcome these problems, we use RESTART/splitting technique to increase the occurrence of these rare events in an acceptable simulation time. This approach exploits the idea to copy equivalent versions of the basic process in the case when it hits predefined thresholds. It allows to increase the number of reaching of the full capacity (implying loss of a customer) considerably.

2. Description of the model

We consider a general single-server $GI/G/1$ -type queueing system with FIFO service discipline. An important feature of the system is that the n th customer is described by two parameters, service time S_n and its *volume* v_n , $n \geq 0$. The pairs $\{S_n, v_n\}$ are assumed to be independent identically distributed (i.i.d.), while, for given n , a dependence between S_n and v_n may exist. In particular, the volumes $\{v_i\}$ are i.i.d. with generic element v , and the service times $\{S_n\}$ are i.i.d. as well, with generic element S . The arrival instants $\{t_n\}$ form the i.i.d (renewal) sequence of the interarrival times $\tau_n = t_{n+1} - t_n$, $n \geq 0$ ($t_0 = 0$) with rate $\lambda := 1/E\tau \in (0, \infty)$ and generic element τ . Because the system capacity is finite, we need not inequality $ES < E\tau$ between the means. In this regard we note that in the system with random volume customers both queue size and the accumulated volume processes are tight.

Denote $V(t)$ the accumulated volume at instant t and let the constant M be an upper bound of $V(t)$ (system capacity) [3]. Note that the buffer for the *number of customers* waiting in the queue is assumed to be infinite. Thus, in this system customer n is lost if and only if

$$V(t_n^-) + v_n > M.$$

Let v_{lost} be a stationary volume of the lost customer. The key observation is that the loss of a customer can be treated as a crossing of the level M by a renewal process generated by the i.i.d. volumes $\{v_k\}$ [4, 5]. Under this interpretation, the volume v_{lost} corresponds to a renewal interval covering the "time instant" M in this renewal process. This gives the following approximations for the two moments:

$$Ev_{lost} = \frac{Ev^2}{Ev}, \quad Ev_{lost}^2 = \frac{Ev^3}{Ev}. \quad (1)$$

In this work we verify the accuracy of these formulae.

3. Simulation

In the experiments, we verify approximation (1) using regenerative simulation and accelerated estimation. As it has been said above, we apply the RESTART/splitting technique to reduce simulation time and, by this, to speed-up estimation of the moments of the lost volume [6, 7]. Unlike classical splitting method, the RESTART (REpetitive Simulation Trials After Reaching Thresholds) method assumes to interrupt a trajectory of the process provided the trajectory falls below a certain level. It allows radically accelerate simulation time comparing with classical splitting in which we wait until each trajectory reaches initial level 0. In our setting, a rare event is the customer loss, and we create a few independent copies

of the accumulated volume process $V(t)$ each time when it reaches one of the (equidistant) thresholds $0 < x_1 < x_2 < \dots < x_n < M$. Moreover, we perform confidence regenerative estimation of the moments of the volume of lost customers.

4. Conclusions

We consider a single-server system in which each customer (besides service time) has a random volume, and the total volume is upper bounded by a finite volume capacity. We apply regenerative simulation in combination with RESTART/splitting technique to speed-up simulation and check the accuracy of the approximation of the moments of the lost customer volume.

Acknowledgments

Research is supported by Russian Foundation for Basic Research, projects 15-07-02341, 15-07-02354, 15-07-02360, and the Program of strategy development of Petrozavodsk State University.

References

1. Tikhonenko O. M. The problem of determination of the summarized messages volume in queueing systems and its applications // *Journal of Information Processing and Cybernetics*. — 1987. — Vol. 32, no. 7. — P. 339–352.
2. Tikhonenko O. M. Queuing system with processor sharing and limited resources // *Automation and Remote Control*. — 2010. — Vol. 71, no. 5. — P. 803–815.
3. Morozov E., Nekrasova R., Potakhina L., Tikhonenko O. Asymptotic analysis of queueing systems with finite buffer space // *Communications in Computer and Information Science*. — 2014. — Vol. 431. — P. 223–232.
4. Asmussen S. *Applied Probability and Queues*. — 2 edition. — Springer-Verlag, NY, 2003.
5. Feller W. *An introduction to probability theory and its applications*. — Wiley, 1971.
6. Garvels M., Kroese D. A comparison of restart implementations // *Proceedings of the 1998 Winter Simulation Conference*. — 1998. — P. 601–608.
7. Kahn H., Harris T. E. Estimation of particle transmission by random sampling // *National Bureau of Standards Applied Mathematics Series*. — 1951.

UDC 004.4

On regenerative envelopes for high performance cluster simulation

E. Morozov^{*†}, I. Peshkova^{*}, A. Rumyantsev^{*†}

^{*} *Petrozavodsk State University*

Lenin str. 33, Petrozavodsk, 185910, Russia

[†] *Institute of Applied Mathematical Research*

of the Karelian research Centre of R.A.S.

Pushkinskaya str. 11, Petrozavodsk, 185910, Russia

Abstract. A regenerative envelopes method for simulation of high performance cluster presented

Keywords: regenerative envelopes, high performance cluster, queue size estimation.

Consider the model of a high performance cluster (HPC) Σ with FCFS service discipline which has m servers working in parallel. Denote by $\tau_n = t_{n+1} - t_n$ the interarrival times (with rate λ), where t_n is the arrival of the n th customer, $n \geq 1$. Customer i requires N_i servers simultaneously for service time S_i (with rate μ) and with distribution $\{p_k := P(N_i = k)\}$. The following Kiefer–Wolfowitz-type recursion describes the dynamics of the workload of the HPC

$$W_{i+1} = R\left(\overbrace{W_{i,N_i} + S_i - \tau_i, \dots, W_{i,N_i} + S_i - \tau_i}^{N_i \text{ components}}, W_{i,N_i+1} - \tau_i, \dots, W_{i,m} - \tau_i\right)^+, \quad (1)$$

where operator R put the components of the vector in an increasing order and $W_{i,j}$ is the remaining work allocated to the j th most busy server at the arrival epoch of customer i . In other words, customer i occupies N_i the least busy servers and, since the servers are stochastically equivalent, we do not take care on the numbers of servers caused by the operator R . Note that if $N = 1$ w.p. 1, then (1) defines the workload vector for a classical GI/G/m multiserver system. Denote $\nu_n(Q_n)$ the number of customers (queue size) at instant t_n^- in the cluster model. The stationary versions ν, Q exist if and only if the following stability criterion holds [8]

$$\frac{\lambda}{\mu} C(m) < 1, \quad (2)$$

where

$$C(m) := \sum_{i=1}^m \frac{1}{i} \sum_{j=i}^m p_j^{*i} \sum_{t=m-j+1}^m p_t, \quad (3)$$

and p_j^{*i} is the i -convolution of p_j with itself. For a 2-server model, the criterion (2) becomes [2]

$$\frac{\lambda}{\mu} \left(1 - \frac{p_1^2}{2} \right) < 1.$$

In the regenerative envelopes method, we construct two queueing systems $\bar{\Sigma}$ (majorant) and $\underline{\Sigma}$ (minorant), with the same (by coupling) input sequence $\{\tau_i\}$ as in Σ , and appropriately enlarged (shortened) service times $\{\bar{S}_i\}$ ($\{\underline{S}_i\}$), $i \geq 1$. The corresponding variables in the system $\bar{\Sigma}$ ($\underline{\Sigma}$) we endow with overline (underline). These transformations occur at the epochs when each system hits the corresponding fixed compact set, implying regeneration. In turn it allows regenerative confidence estimation of the system [4–6] even if the original system is not regenerative. (More on this method see in [9].) The procedure is based on the following monotonicity result which holds true for the HPC model as well as for a classical $GI/G/m$ model [1]. Assume the initial workloads ordered as $\underline{W}_1 \leq W_1 \leq \bar{W}_1$, and the service times are ordered as $\underline{S}_i \leq S_i \leq \bar{S}_i$, $i \geq 1$. Then

$$\underline{\nu}_i \leq \nu_i \leq \bar{\nu}_i, \underline{Q}_i \leq Q_i \leq \bar{Q}_i, i \geq 1.$$

To define the compact set, we use the following two methods:

1) Fix integer Q_0 (the number of customers in the queue) and N_0 (the number of servers required by the customer at the head of the queue) and constants $0 \leq a \leq b < \infty$, and define

$$\bar{\beta}_{n+1} = \inf \left\{ k > \bar{\beta}_n : \bar{Q}_k = Q_0 > 0, N_{k-Q_0} = N_0, \bar{S}_i(k) \in (a, b), i \in \bar{\mathcal{M}}_k, \bar{\mathcal{M}}_{\bar{\beta}_n} \cap \bar{\mathcal{M}}_k = \emptyset \right\}, n \geq 0,$$

where $\bar{\mathcal{M}}_n = \{i : t_i \leq t_n < \bar{z}_i\}$ is the set of the customers served in the system $\bar{\Sigma}$ at instant t_n , $\bar{S}_i(n)$ is the remaining service time of customer i at instant t_n and \bar{z}_n is the departure instant of n th customer.

2) Fix integer m_0 (the number of busy servers, if the queue is empty) and define

$$\bar{\beta}_{n+1} = \inf \left\{ k > \bar{\beta}_n : \bar{Q}_k = 0, \sum_{i \in \bar{\mathcal{M}}_k} N_i = m_0 \leq m, \bar{S}_i(k) \in (a, b), i \in \bar{\mathcal{M}}_k, \bar{\mathcal{M}}_{\bar{\beta}_n} \cap \bar{\mathcal{M}}_k = \emptyset \right\}, n \geq 0.$$

We define the moments $\bar{\beta}_n$, $n \geq 0$ in a similar way (with possibly other values a, b, Q_0, N_0, m_0). When the corresponding system hits the compact

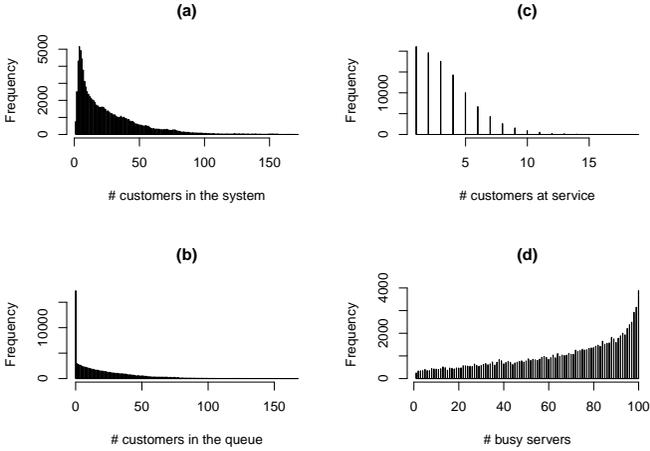


Figure 1. Histograms of the performance measures for $m = 100$ servers, Poisson arrivals with rate $\lambda = 1$, exponential service time with rate $\mu = 0.3$ and $p_i = p_0/i, i = 1, \dots, s$: (a) customers in the system, (b) customers in the queue, (c) customers at service, (d) busy servers.

set, we replace the values $S_i(k), i \in \mathcal{M}_k$ by b (in $\bar{\Sigma}$) and by a (in $\underline{\Sigma}$). It is easy to show that the constructed moments $\{\bar{\beta}_n\}, \{\underline{\beta}_n\}$ are regeneration instants of the corresponding basic processes, and moreover, these replacements keep the monotonicity of the performance measures of the interest. We stress that, at the regeneration instant, all the remaining service times of the customers being served become equal to b , and can be treated a single *large* customer occupying m_0 servers. Note that for a classical $GI/G/m$ system the cases 1) and 2) coincide, since

$$N = 1 \text{ w.p. } 1, \sum_{i \in \mathcal{M}_k} N_i = \nu_k \text{ and } Q_k = \max(0, \nu_k - m).$$

The choice of constants a, b, Q_0, N_0, m_0 leads to a trade-off between the frequency of the regenerations and the divergence of the upper, original and lower models. In this regard we mention that the (well-known in classical $GI/G/m$ system, see [1]) mode of the stationary number of the customers in the system $\nu = \lceil \lambda/\mu \rceil$ surprisingly extends to the HPC model. However, there is no easy relation between the number of customers in the system ν and the number of customers at service (in the queue), as well as the number of the busy servers. To illustrate this, we perform a numerical experiment. We take $m = 100$ and generate a sequence of

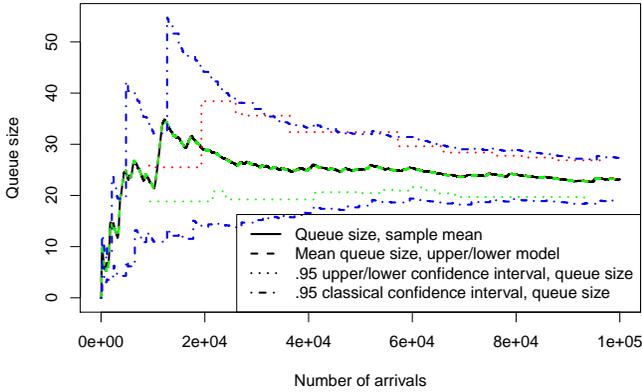


Figure 2. Confidence intervals for the system with $m = 100$ servers, Poisson arrivals with rate $\lambda = 1$, exponential service time with rate $\mu = 0.3$ and $p_i = p_0/i, i = 1, \dots, s$.

$n = 10^5$ Poisson arrivals (with rate $\lambda = 1$). We set $p_i = p_0/i, i = 1, \dots, m$, where $p_0 = (p_1 + \dots + p_m)^{-1}$. (It is Zipf’s law, the discrete analogue of a heavy-tailed Pareto law.) Then we select $\mu = 0.9\lambda C(100)$, with $C(100)$ defined in (3), which provided $\lceil \lambda/\mu \rceil = 4$. We calculate the following performance measures: the number of customers in the system, the number of customers in the queue, the number of customers being served, the number of busy servers. The histograms of the results are shown on Fig. 1. It is seen that the most frequent number of customers in the system equals 4, and it corresponds to our conjecture. However, there is mostly only 1 customer at service. There is no queue in the most of cases (approx. 15%), which is related to geometrically decreasing probability $P(\nu = k)$ for k large. Note also that it is more probably that more than $m/2$ servers are busy, which is a well-known effect for a heavy-tailed distribution. Thus, we select $Q_0 = 0$ both for the upper and for the lower models, and set $m_0 = 100$, as the most frequent value. Next, we select (a, b) as $(0, 2.4)$ for the upper system, and $(2.2, \infty)$ for the lower system. We apply regenerative confidence estimations for the regenerative envelopes, and for the original system as well, since the original system possesses classical regenerations (when $\nu_i = 0$). The results of the experiment are shown on Fig. 2. It is seen that the regenerative envelop method slightly outperforms the classical confidence interval.

Acknowledgment

This work is supported by Russian Foundation for Basic research, projects No 15-07-02341, 15-07-02354, 15-07-02360, 15-29-07974, 16-07-00622 and by the Program of strategic development of Petrozavodsk State University.

References

1. Asmussen S.: Applied Probability and Queues. Wiley, N.Y. (1987)
2. Chakravarthy S., Karatza H.: Two-server parallel system with pure space sharing and markovian arrivals. *Computers & Operations Research*. 40, No. 1, 510–519 (2013)
3. CRAN - Package hpcwld, <http://cran.r-project.org/web/packages/hpcwld/index.html>
4. Glynn P.: Some topics in regenerative steady-state simulation. *Acta Appl. Math.* 34, 225–236 (1994)
5. Glynn P., Iglehart D.: Conditions for the applicability of the regenerative method. *Management Sci.* 39, 1108–1111 (1993)
6. Sigman K., Wolff R. W.: A review of regenerative processes. *SIAM Review*. 35, 269–288 (1993)
7. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>
8. Rumyantsev A., Morozov E.: Stability criterion of a multiserver model with simultaneous service. *Ann Oper Res.* 1–11 (2015)
9. Morozov E., Nekrasova R., Peshkova I., and Rumyantsev A. A Regeneration-Based Estimation of High Performance Multiserver Systems // In *Proceedings of Computer Networks: 23rd International Conference, CN 2016, Brunow, Poland, June 14–17, 2016*, / Cham: Springer International Publishing, 2016, pp. 271–282.

УДК 519.872

Асимптотический анализ RQ-системы M|M|1 с нетерпеливыми заявками в условии долгой терпеливости

А. А. Назаров, Е. А. Фёдорова

*Томский государственный университет,
пр. Ленина, 36, Томск, Россия, 634050*

Аннотация. В работе рассматривается однолинейная RQ-система с нетерпеливыми заявками M/M/1, время задержки заявок на орбите и время терпения заявок распределены экспоненциально. Исследуется процесс, описывающий число заявок в источнике повторных вызовов (орбите) в системе. Для распределения вероятностей состояний системы была составлена система дифференциальных уравнений Колмогорова, которая была записана в стационарном режиме. Для решения полученной системы уравнений авторами предлагается метод асимптотического анализа в условии долгой терпеливости заявок. В результате исследования были получены асимптотики первого и второго порядков. Таким образом, было показано, что асимптотическая характеристическая функция числа заявок в источнике повторных вызовов имеет вид нормального распределения. Было проведено численное сравнение асимптотического и точного распределений (полученного с помощью численного алгоритма), на основе чего, был сделан вывод об области применимости асимптотического метода.

Ключевые слова: RQ-система, нетерпеливые заявки, асимптотический анализ.

1. Введение

Системы массового обслуживания с повторными вызовами или RQ-системы (Retrial Queueing System) – математические модели, широко применяемые для анализа и оптимизации различных телекоммуникационных систем, сетей мобильной связи, call-центров и др [1–4]. Характерной чертой таких систем является наличие повторных обращений заявок к обслуживающему прибору спустя некоторое случайное время после неудачной попытки обслуживания. Такие ситуации могут быть вызваны не только отсутствием свободных серверов в моменты поступления заявок, но некоторыми техническими причинами. Детальное описание RQ-систем и обзор результатов их исследования представлены в монографиях [5, 6].

Существует достаточно большое число работ, посвященное и RQ-системам с нетерпеливыми заявками. Однако в большинстве из них нетерпеливость понимается в том смысле, что заявка с определенной вероятностью уходит из ИПВ после неудачной попытки обслуживания [6]. В данной работе предполагается, что терпеливость заявки описывается по экспоненциальному закону (впоследствии, будет совершен переход к произвольному закону).

2. Описание модели

Рассмотрим (рис. 1) однолинейную RQ-систему, на вход которой поступает простейший поток заявок с параметром λ , время обслуживания каждой заявки распределено по экспоненциальному закону с параметром μ . Если поступившая заявка застаёт прибор свободным, то оно занимает его для обслуживания. Если прибор занят, то заявка переходит в ИПВ, где осуществляет случайную задержку, продолжительность которой имеет экспоненциальное распределение с параметром σ . Из ИПВ после случайной задержки заявка вновь обращается обслуживающему прибору с повторной попыткой его захвата. Если прибор свободен, то заявка из ИПВ занимает его для обслуживания, в противном случае заявка мгновенно возвращается в источник повторных вызовов для реализации следующей задержки. Из ИПВ после случайного времени, имеющего экспоненциальное распределение с параметром α , заявка покидает систему.

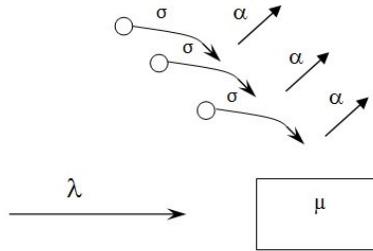


Рис. 1. RQ-система M|M|1 с нетерпеливыми заявками

Пусть $i(t)$ – число заявок в ИПВ, а $k(t)$ – определяет состояние прибора следующим образом:

$$k(t) = \begin{cases} 0, & \text{прибор свободен,} \\ 1, & \text{прибор занят.} \end{cases}$$

Обозначим $P(k, i, t) = P\{k(t) = k, i(t) = i\}$ – вероятность того, что прибор в момент времени t находится в состоянии k и в источнике повторных вызовов находится i заявок. Очевидно, что процесс $\{k(t), i(t)\}$ изменения состояний данной системы во времени является марковским. Для распределения вероятностей $P(k, i, t)$ состояний рассматриваемой RQ-системы составим систему дифференциальных

уравнений Колмогорова:

$$\begin{cases} \frac{\partial P_0(i, t)}{\partial t} = -(\lambda + i\sigma + i\alpha)P_0(i, t) + \mu P_1(i, t) + \alpha(i+1)P_0(i+1, t), \\ \frac{\partial P_1(i, t)}{\partial t} = -(\lambda + \mu + i\alpha)P_1(i, t) + \lambda P_0(i, t) + \sigma(i+1)P_0(i+1, t) + \\ + \lambda P_1(i-1, t) + \alpha(i+1)P_1(i+1, t). \end{cases} \quad (1)$$

Обозначим $P_k(i) = \lim_{t \rightarrow \infty} P_k(i, t)$. Тогда в стационарном режиме система (1) примет вид:

$$\begin{cases} -(\lambda + i\sigma + i\alpha)P_0(i) + \mu P_1(i) + \alpha(i+1)P_0(i+1) = 0, \\ -(\lambda + \mu + i\alpha)P_1(i) + \lambda P_0(i) + \sigma(i+1)P_0(i+1) \\ + \lambda P_1(i-1) + \alpha(i+1)P_1(i+1) = 0, \end{cases} \quad (2)$$

Перейдем к частичным характеристическим функциям:

$$H_k(u) = \sum_i e^{ju_i} P_k(i),$$

где $j = \sqrt{-1}$ - мнимая единица.

Система уравнений (2) переписется в виде:

$$\begin{cases} j(\sigma + \alpha(1 - e^{-ju}))H'_0(u) + \mu H_1(u) - \lambda H_0(u) = 0, \\ j(\sigma + \alpha)H'_0(u) + \lambda e^{ju} H_1(u) + j\alpha H'_1(u) = 0. \end{cases} \quad (3)$$

Полученную систему будем решать методом асимптотического анализа в условии долгой терпеливости заявок (long time of patience), то есть при $\alpha \rightarrow 0$.

3. Первая асимптотика

Введем обозначения $\alpha = \varepsilon$, $u = w\varepsilon$, $H_0(u) = \varepsilon F_0(w, \varepsilon)$, $H_1(u) = F_1(w, \varepsilon)$, где $\varepsilon \rightarrow 0$ - бесконечно малая величина.

Тогда система (3) примет вид:

$$\begin{cases} j(\sigma + \varepsilon(1 - e^{-jw\varepsilon})) \frac{\partial F_0(w, \varepsilon)}{\partial w} + \mu F_1(w, \varepsilon) - \lambda \varepsilon F_0(w, \varepsilon) = 0, \\ j(\sigma + \varepsilon) \frac{\partial F_0(w, \varepsilon)}{\partial w} + \lambda e^{jw\varepsilon} F_1(w, \varepsilon) + j \frac{\partial F_1(w, \varepsilon)}{\partial w} = 0. \end{cases} \quad (4)$$

Совершив в (4) предельный переход $\varepsilon \rightarrow 0$ и обозначив $F_k(w) = \lim_{\varepsilon \rightarrow 0} F_k(w, \varepsilon)$, получим:

$$\begin{cases} j\sigma F_0'(w) + \mu F_1(w) = 0, \\ j\sigma F_0'(w) + \lambda F_1(w) + jF_1'(w) = 0. \end{cases} \quad (5)$$

В результате преобразований из (5) имеем:

$$F_1(w) = R_1 \exp \{jw(\lambda - \mu)\}.$$

Допредельная характеристическая функция равна:

$$h(u) = H_1(u) + H_0(u) = F_1(u/\varepsilon) + O(\varepsilon).$$

Асимптотикой первого порядка будем называть функцию:

$$h^{(1)}(u) = F_1(u/\varepsilon).$$

Нетрудно показать, что $h^{(1)}(u) = \exp \left\{ ju \frac{\lambda - \mu}{\alpha} \right\}$.

4. Вторая асимптотика

Найдем асимптотику второго порядка, для этого в систему (3) подставим характеристическую функцию вида:

$$H_k(u) = H_k^{(2)}(u) \exp \left\{ ju \frac{\lambda - \mu}{\alpha} \right\}.$$

Получим следующую систему уравнений:

$$\begin{cases} j(\sigma + \alpha(1 - e^{-ju}))(H_0^{(2)}(u))' + \mu H_1^{(2)}(u) - \\ - \left[\frac{\lambda - \mu}{\alpha} \sigma + (\lambda - \mu)(1 - e^{-ju}) + \lambda \right] H_0^{(2)}(u) = 0, \\ j\alpha(H_1^{(2)}(u))' + j(\sigma + \alpha)(H_0^{(2)}(u))' + \\ + [\lambda e^{ju} - \lambda + \mu] H_1^{(2)}(u) - \frac{\lambda - \mu}{\alpha} (\sigma + \alpha) H_0^{(2)}(u) = 0. \end{cases} \quad (6)$$

Введем обозначения:

$$\alpha = \varepsilon^2, \quad u = w\varepsilon, \quad H_0^{(2)}(u) = \varepsilon^2 F_0^{(2)}(w, \varepsilon), \quad H_1^{(2)}(u) = F_1^{(2)}(w, \varepsilon). \quad (7)$$

Будем использовать следующие разложения функций:

$$F_k^{(2)}(w, \varepsilon) = F_k^{(2)}(w) + jw\varepsilon f_k(w) + O(\varepsilon^2). \quad (8)$$

Подставляя (7) и (8) в систему (6), получим следующие уравнения:

$$\begin{cases} j\sigma\varepsilon(F_0^{(2)}(w))' + \mu F_1^{(2)}(w) + \mu jw\varepsilon f_1(w) - (\lambda - \mu)\sigma F_0^{(2)}(w) - \\ - (\lambda - \mu)\sigma jw\varepsilon f_0(w) = O(\varepsilon^2), \\ j\varepsilon(F_1^{(2)}(w))' + j\sigma\varepsilon F_0^{(2)}(w) + [\lambda(1 + jw\varepsilon) - \lambda + \mu] F_1^{(2)}(w) + \\ + \mu jw\varepsilon f_1(w) - (\lambda - \mu)\sigma F_0^{(2)}(w) - jw\varepsilon(\lambda - \mu)\sigma f_0(w) = O(\varepsilon^2). \end{cases} \quad (9)$$

Совершим в (9) предельный переход $\varepsilon \rightarrow 0$:

$$\mu F_1^{(2)}(w) - (\lambda - \mu)\sigma F_0^{(2)}(w) = 0. \quad (10)$$

Подставим (10) в (9), сократим на общий множитель. Совершив переход $\varepsilon \rightarrow 0$, нетрудно получить следующее:

$$jF_1^{(2)}(w) + \lambda jwF_1^{(2)}(w) = 0. \quad (11)$$

Решая (10) и (11), получим:

$$F_1^{(2)}(w) = R_2 \exp\left\{-\lambda \frac{w^2}{2}\right\}.$$

Возвращаясь к заменам (7), имеем:

$$H_1^{(2)}(u) = F_1^{(2)}(u/\varepsilon) = \exp\left\{-\lambda \frac{u^2}{2\alpha}\right\}.$$

Асимптотикой второго порядка будем называть функцию:

$$h^{(2)}(u) = H_1^{(2)}(u)h_1(u).$$

Тогда $h^{(2)}(u)$ равна:

$$h^{(2)}(u) = \exp\left\{ju \frac{\lambda - \mu}{\alpha} - \lambda \frac{u^2}{2\alpha}\right\}.$$

Таким образом, мы получили, что асимптотическая характеристическая функция числа заявок в источнике повторных вызовов имеет нормальное распределение с математическим ожиданием $(\lambda - \mu)/\alpha$ и дисперсией λ/α .

5. Численный анализ результатов

Исследуем область применимости предлагаемого асимптотического метода на основе численных экспериментов. Систему (2) можно решить численно с помощью рекуррентного алгоритма (предварительно выполнив некоторые преобразования) или представив систему линейных уравнений в виде матричного уравнения решая его с помощью любого известного метода.

В качестве критерия близости распределений в работе предлагается использовать расстояние Колмогорова, которое вычисляется следующим образом:

$$\Delta = \max_{0 \leq i \leq I_{max}} \left| \sum_{n=0}^i D(n) - \sum_{n=0}^i P(n) \right|.$$

где $P(n)$ – распределение вероятностей числа заявок в источнике повторных вызовов, полученное численным алгоритмом, а $D(n)$ – асимптотическое распределение.

В качестве примера продемонстрируем зависимость расстояния Колмогорова от значений α при следующих значениях параметров системы: $\sigma = 10$, $\mu = 1$ (табл. 1).

Таблица 1

Расстояние Колмогорова между асимптотическим и точным распределениями

α	1	0.5	0.1	0.05	0.01	0.005
$\lambda = 1.5$	0.163	0.125	0.024	0.007	0.005	0.004
$\lambda = 3$	0.076	0.037	0.008	0.006	0.003	0.002
$\lambda = 5$	0.039	0.020	0.007	0.006	0.002	0.001
$\lambda = 10$	0.020	0.014	0.006	0.004	0.002	0.001

6. Заключение

Таким образом, в работе показано, что в условии долгой терпимости заявок характеристическая функция числа заявок в источнике повторных вызовов может аппроксимироваться нормальным распределением с математическим ожиданием $(\lambda - \mu)/\alpha$ и дисперсией λ/α . Принимая допустимым значение расстояния Колмогорова между распределениями $\Delta \leq 0.03$, можно сделать вывод, что предлагаемый в работе метод асимптотического анализа применим уже при $\alpha \leq 0.1$.

Благодарности

Работа выполнена в рамках проекта Министерства образования и науки РФ № 1.511.2014/К.

Литература

1. *Wilkinson R. I.* Theories for toll traffic engineering in the USA // The Bell System Technical Journal, 1956. — Vol. 35, no. 2. — P. 421–507.
2. *Cohen J.W.* Basic problems of telephone traffic and the influence of repeated calls // Philips Telecommunication Review, 1957. — Vol. 18, no. 2. — P. 49–100.
3. *Gosztony G.* Repeated call attempts and their effect on traffic engineering // Budavox Telecommunication Review, 1976. — Vol. 2. — P. 16–26.
4. *Kuznetsov D. Yu., Nazarov A.A.* Analysis of non-Markovian models of communication networks with adaptive protocols of multiple random access // *Avtomatika i Telemekhanika*, 2001. — Vol. 5. — P. 124–146.
5. *Artalejo J.R., Gómez-Corral A.* Retrial Queueing Systems. A Computational Approach. — Springer, 2008.
6. *Falin G.I., Templeton J.G.C.* Retrial queues. — Chapman & Hall London, 1997.

UDC 519.872

Asymptotic analysis of retrial queue M/M/1 with impatient calls under the long patience time condition

A. Nazarov, E. Fedorova

*Tomsk state university
Lenina str., 36, Tomsk, 634050 Russia*

In the paper the retrial queueing system M/M/1 with impatient calls is studied. The delay time of each call in the orbit and the patience of customers are exponentially distributed. The problem is to find the probability distribution of the number of calls in the orbit in the system. The direct system of Kolmogorov differential equations for the system states probability distribution is written in the stationary state. The method of asymptotic analysis under the condition of long time of calls patience is offered for the system of differential equations investigating. In the paper the first and the second order asymptotic functions are obtained. In this way, it is proved that the asymptotic characteristic function under the condition of long time of patience has the normal distribution. The numerical comparison of asymptotic and exact distributions (obtained numerically) is performed and the area of the method application is defined.

Keywords: retrial queueing system, impatient calls, asymptotic analysis.

UDC 519.2

Inventory management system with On/Off control and phase-type distribution of purchases quantity

A. A. Nazarov, V. I. Broner

*Department of probability theory and mathematical statistics
National Research Tomsk State University
Lenin ave. 36, Tomsk, 634050, Russia*

Abstract. The purpose of this paper is to study mathematical model of an inventory management system with On/Off control. We consider the case in which input flow of product is continuous with fixed rate. Demand occurs according to a Poisson process with constant intensity and quantity of purchase have phase-type Distribution. We find an explicit form for a stationary probability density function of inventory level.

Keywords: Inventory management, On/Off control, mathematical modelling, phase-type distribution.

1. Problem statement

Inventory control models under various conditions have been studied intensively in the last century, for example, Single-period and Newsvendor problem are widely-known [1–5]. This models used to analyse systems with perishable products in airlines, fashion industries and other fields. In this paper we propose stochastic mathematical model of inventory management with on/off control.

Consider an product which is demanded and the product flow be continuous with fixed rate $\nu = 1$ (Fig. 1).

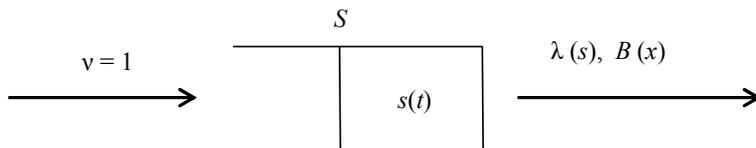


Figure 1. Inventory management system

Let $s(t)$ be inventory level at time t . The demand for the occurs according to a Poisson process with piecewise constant intensity $\lambda(s)$

$$\lambda(s) = \begin{cases} \lambda_1, & s < S, \\ \lambda_2, & s \geq S, \end{cases}$$

where S is the threshold inventory level of $s(t)$.

The values of purchases are independent and identically distributed random variables having Phase-type distribution

$$B(x) = 1 - \beta e^{\mathbf{G}x\mathbf{E}},$$

where $\beta_k > 0$, \mathbf{G} is subgenerator matrix Markov chain that determines the Phase-type distribution and

$$\beta\mathbf{E} = 1.$$

In this paper, it is assumed that the process $s(t)$ can take the values $s(t) < 0$. In this situation, the customer waits for the required amount of product.

Condition for the existence of a stationary distribution has form

$$\lambda_1 b < 1 < \lambda_2 b,$$

where b is the first moment of the probability distribution (1).

Clearly, if condition if $\lambda_1 < 1/b < \lambda_2$ and $s(t) < S$ is satisfied then the stock level increases in the mean. Otherwise condition $s(t) \geq S$ means that the stock level decreases in the mean.

based on the problem statement, we conclude that $s(t)$ is a Markovian process with continuous time t and continuous state space $-\infty < s < \infty$.

We denote by

$$P(s, t) = \frac{\partial P \{s(t) < s\}}{\partial s}$$

the stationary probability density function of stock level.

We can derive the equation

$$P(s + \Delta t) = P(s)(1 - \lambda(s)\Delta t) + \Delta t \int_0^\infty \lambda(s+x)P(s+x)dB(x) + o(\Delta t).$$

We obtain Kolmogorov equation for the distribution $P(s)$

$$P'(s) + \lambda(s)P(s) = \int_0^\infty \lambda(s+x)P(s+x)dB(x),$$

where the boundary conditions have following form

$$P(-\infty) = P(\infty) = 0.$$

Let us find a solution $P(s)$ of the Kolmogorov equation (1) in an explicit form, which is satisfied the boundary conditions (1).

Let us denote

$$P(s) = \begin{cases} P_1(s), & s < S, \\ P_2(s), & s > S. \end{cases}$$

Therefore, we can rewrite the equation (1) as two equations

$$P_2'(s) + \lambda_2 P_2(s) = \lambda_2 \int_0^\infty P_2(s+x)dB(x), \quad s > S,$$

and

$$P_1'(s) + \lambda_1 P_1(s) = \lambda_1 \int_0^{S-s} P_1(s+x)dB(x) + \lambda_2 \int_{S-s}^\infty P_2(s+x)dB(x), \quad s < S.$$

Now we find solutions of equations (1) and (1), that satisfy the boundary conditions

$$P_1(-\infty) = 0, P_2(\infty) = 0.$$

2. The solution $P_2(s)$ of equation (1)

Solution $P_2(s), s > S$ of equation (1) has to be sought in the form

$$P_2(s) = Ce^{-\gamma(s-S)}, \quad s > S.$$

using substitution (2) into (1)), we derive the equation

$$\lambda_2 - \gamma = \lambda_2 \int_0^\infty e^{-\gamma x} dB(x).$$

Obviously, that equation (2) has extraneous zero root $\gamma = 0$, because we have the boundary condition (1) $P_2(\infty) = 0$.

It is easy to see that unique positive root $\gamma > 0$ of equation (2) exists for any distribution function $B(x)$ under the condition (1) $\lambda_2 b > 1$, consequently the solution of the equation (1) is a function (1) defined with multiplicative constant C accuracy, which value will be find later.

3. The solution $P_1(s)$ of equation (1)

Taking into account (2), we can rewrite equation (1) in the form

$$P_1'(s) + \lambda_1 P_1(s) = \lambda_1 \int_0^{S-s} P_1(s+x) dB(x) + \lambda_2 C e^{-\gamma(s-S)} \int_{S-s}^{\infty} e^{-\gamma x} dB(x).$$

Using (1), we can find the integral on the right side of the equation (3)

$$\begin{aligned} \int_{S-s}^{\infty} e^{-\gamma x} dB(x) &= - \int_{S-s}^{\infty} e^{-\gamma x} \beta e^{\mathbf{G}x} \mathbf{G} \mathbf{E} dx = - \int_{S-s}^{\infty} \beta e^{(\mathbf{G}-\gamma \mathbf{I})x} \mathbf{G} \mathbf{E} dx = \\ &= \beta e^{(\mathbf{G}-\gamma \mathbf{I})(S-s)} (\mathbf{G} - \gamma \mathbf{I})^{-1} \mathbf{G} \mathbf{E}, \end{aligned}$$

then (3) can be written as follows

$$P_1'(s) + \lambda_1 P_1(s) = \lambda_1 \int_0^{S-s} P_1(s+x) dB(x) + \lambda_2 C \beta e^{(\mathbf{G}-\gamma \mathbf{I})(S-s)} (\mathbf{G} - \gamma \mathbf{I})^{-1} \mathbf{G} \mathbf{E}.$$

Substituting the expression (1) for distribution functions $B(x)$ in this equation, we obtain the following equation

$$P_1'(s) + \lambda_1 P_1(s) = \beta \left(\lambda_1 \int_0^{S-s} P_1(s+x) e^{\mathbf{G}x} dx - \lambda_2 C e^{(\mathbf{G}-\gamma \mathbf{I})(S-s)} (\mathbf{G} - \gamma \mathbf{I})^{-1} \right) \mathbf{G} \mathbf{E}.$$

Theorem 1 *If the equation*

$$z + \lambda_1 = \lambda_1 \beta (\mathbf{G} + z \mathbf{I})^{-1} \mathbf{G} \mathbf{E}$$

has n simple roots with positive real parts, then solution $P_1(s)$ of equation (3) has form

$$P_1(s) = C \sum_{\nu=1}^n x_{\nu} e^{z_{\nu}(s-S)}, s < S,$$

where $z = z_{\nu}$, $\nu = \overline{1, n}$ is a nonzero roots of equation (1), $x_{\nu}, \nu = \overline{1, n}$ are solutions to a system of equations

$$\left(\lambda_1 \sum_{\nu=1}^n x_{\nu} (\mathbf{G} + z_{\nu} \mathbf{I})^{-1} - \lambda_2 (\mathbf{G} - \gamma \mathbf{I})^{-1} \right) \mathbf{G} \mathbf{E} = 0,$$

normalizing constant C is determined by the equation

$$C = \left(\frac{1}{\gamma} + \sum_{\nu=1}^n \frac{x_\nu}{z_\nu} \right)^{-1}.$$

Proof. Solution $P_1(s)$ of the equation (3) will be find in the form (1). Substituting (1) into (3) we obtain the equation

$$\begin{aligned} & \sum_{\nu=1}^n x_\nu e^{z_\nu(s-S)} \left\{ z_\nu + \lambda_1 - \lambda_1 \beta (\mathbf{G} + z_\nu \mathbf{I})^{-1} \mathbf{G} \mathbf{E} \right\} = \\ & = \beta e^{\mathbf{G}(S-s)} \left(\lambda_1 \sum_{\nu=1}^n x_\nu (\mathbf{G} + z_\nu \mathbf{I})^{-1} - \lambda_2 (\mathbf{G} - \gamma \mathbf{I})^{-1} \right) \mathbf{G} \mathbf{E}. \end{aligned}$$

By equating the coefficients to zero in the linear combination of exponents $e^{z_\nu(s-S)}$ in this expression, we get

$$z_\nu + \lambda_1 = \lambda_1 \beta (\mathbf{G} + z_\nu \mathbf{I})^{-1} \mathbf{G} \mathbf{E}, \quad \nu = \overline{1, n}.$$

Obviously that this expression and (1) have the same form. Consequently, $z_\nu, \nu = \overline{1, n}$ are the roots of the equation (1).

Analogically, we obtain

$$\left(\lambda_1 \sum_{\nu=1}^n x_\nu (\mathbf{G} + z_\nu \mathbf{I})^{-1} - \lambda_2 (\mathbf{G} - \gamma \mathbf{I})^{-1} \right) \mathbf{G} \mathbf{E} = 0.$$

Using the normalization condition we derive the constant C

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} P(s) ds = \int_{-\infty}^S P_1(s) ds + \int_S^{\infty} P_2(s) ds = \\ &= C \sum_{\nu=1}^n x_\nu \int_{-\infty}^S e^{z_\nu(s-S)} ds + C \int_S^{\infty} e^{-\gamma(s-S)} ds = \\ &= C \sum_{\nu=1}^n x_\nu \int_{-\infty}^0 e^{z_\nu x} dx + C \int_0^{\infty} e^{-\gamma x} dx = C \left\{ \sum_{\nu=1}^n \frac{x_\nu}{z_\nu} + \frac{1}{\gamma} \right\}. \end{aligned}$$

Finally, we get

$$C = \left(\sum_{\nu=1}^n \frac{x_\nu}{z_\nu} + \frac{1}{\gamma} \right)^{-1}.$$

It is easy to see that this expression coincides with (1).

The theorem is proved. ■

Probability density function $P(s)$ of stock-level process has form

$$P(s) = \left(\sum_{\nu=1}^n \frac{x_{\nu}}{z_{\nu}} + \frac{1}{\gamma} \right)^{-1} \cdot \begin{cases} \sum_{\nu=1}^n x_{\nu} e^{z_{\nu}(s-S)}, & s < S, \\ e^{-\gamma(s-S)}, & s > S, \end{cases}$$

where z_{ν} is a nonzero roots of equation (1), γ is unique positive root of equation (2), x_{ν} are solutions equations (1).

4. Numerical results

Let us consider Phase-type distribution of random demand with 3 phases.

For the following values of the parameters $\lambda_1 = 0.8$ and $\lambda_2 = 1.2$, $S = 10$ We found the roots of equations (3) and (1). Thus, the equation (2) has a unique positive solution $\gamma = 0.198$, the equation (1) has three real roots $z_1 = 0.198$, $z_2 = 6.327$, $z_3 = 8.345$. Let us find probability density function of inventory level for the given parameters.

The parameters $x_{\nu}, \nu = \overline{1, n}$ and normalizing constant of distribution (3), have the form

$$x_1 = 1; x_2 = -0.004; x_3 = 0.003, C = 0.99,$$

resulting distribution is shown in Fig. 2.

The explicit expression (3) for the solution $P(s)$ of the equation (1) completely solves the problem of the study of mathematical inventory control model with following restrictions: on/off control and Phase-type distribution of values of product purchases.

Acknowledgments

The work is performed under the state order No. 1.511.2014/K of the Ministry of Education and Science of the Russian Federation.

References

1. Arrow, K.J., Harris, Th.E., Marschak, J.: Optimal Inventory Policy. *Econometrica*, 19 (3), 205–272 (1951)
2. Silver, E.A., Pyke, D.F., Peterson R.: *Inventory Management and Production Planning and Scheduling*. Wiley, New York (1998)
3. Qin, Y., Wang, R., Vakharia, A., Chen, Y., Hanna-Seref, M.: The newsvendor problem: review and directions for future research. *European Journal of Operational Research*, 213, 361–374 (2011)

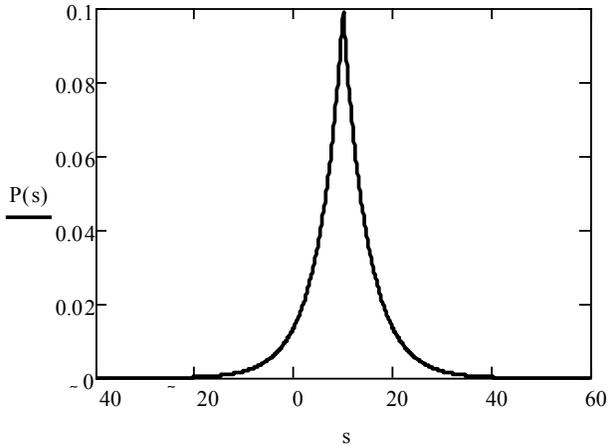


Figure 2. probability density function $P(s)$

4. Kitaeva, A., Subbotina, V., Zmeev, O.: The Newsvendor Problem with Fast Moving Items and a Compound Poisson Price Dependent Demand. In: 15th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2015, Elsevier, (IFAC-PapersOnLine), 48, 1375–1379 (2015)
5. Khouja, M.: The single-period (news-vendor) problem: literature review and suggestions for future research. OMEGA-INT J, 27(5), 537–553 (1999)

UDC 004.4

Asymptotic analysis of M/GI/1 retrial system with conflicts and afterservice

A. A. Nazarov, N. A. Pomortseva

*Tomsk State University
Lenina ave. 36, Tomsk, 634050, Russia*

Abstract. M/GI/1 retrial queueing system with conflicts and afterservice is investigated. The distribution of the request's number in the blocks of the orbit is studied. Throughput value and characteristic function of the number of requests in the blocks of the orbit are obtained for this system.

Keywords: retrial queueing system, conflict, afterservice, asymptotic analysis, long delay.

1. Introduction

A large number of important practical problems arising in connection with the rapid development of information, computer, telecommunication systems can be solved by queueing theory.

In this paper we consider a single-server retrial queueing system with conflicts. A conflict situation suggests that a request, which comes in the system and finds the server busy, and a request under service enter into a conflict, and the both are sent to the orbit. Such models are widely used in real life systems, for example, in computer networks managed by random multiple access protocols.

Falin and Sukharev [1] have analyzed the retrial queueing system with collision, called the queue with double connections. Choi et al. [2] have considered retrial queues with constant retrial rate and collision arising from unslotted CSMA/CD protocol. Krishna Kumar B., Vijayalakshmi G., Krishnamoorthy A., Sadiq Basha S. [3] have analyzed a Markovian single server feedback retrial queue with linear retrial rate and collisions of customers. For this system the joint steady-state probability generating functions of the server state and the orbit length are obtained and some important performance measures of this system are calculated.

In this paper we consider the M/GI/1 retrial system with conflicts of requests and afterservice. For a probability distribution of the system states, we obtain Kolmogorov system of equations, we get the throughput of this queueing system and the characteristic function of the number of requests in the blocks of the orbit.

2. System description and problem statement

Let us consider a single-server retrial queueing system (Figure 1) with a stationary Poisson arrival process of requests with parameter λ . The

service time of the request is a random variable, which has a two-phase distribution with parameters μ_1 for the first phase and μ_2 for the second one. If an arriving request finds the server free, the request occupies the server and gets a service. If the server is busy, the arriving request and the request under service enter into a conflict, and the both are sent to the orbit, which consists of two blocks. In the first block of the orbit, a random delay is performed by the requests, service of which has not been completed at the first phase. In the second block of the orbit, a random delay is performed by the requests, service of which was interrupted by the conflict at the second phase. After a random delay, which has an exponential distribution with parameters σ_1 for the first block and σ_2 for the second block, each request from the k -th block of the orbit turns to the k -th phase of the service, that represents a procedure of requests afterservice.

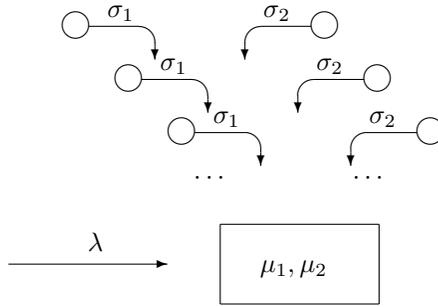


Figure 1. A single-server retrial system

Let $i_1(t)$ be a requests number in the first block of the orbit, $i_2(t)$ be a requests number in the second block of the orbit, and process $k(t)$ defines the state of the server. Our goal is to investigate a three-dimensional random process $\{k(t), i_1(t), i_2(t)\}$.

3. The system of Kolmogorov equations

Denoting

$$P_k(i_1, i_2, t) = P\{k(t) = k, i_1(t) = i_1, i_2(t) = i_2\},$$

$$H_k(u_1, u_2) = \sum_{i_1, i_2=0}^{\infty} \exp\{ju_1i_1 + ju_2i_2\} P_k(i_1, i_2),$$

it is possible to write down the system of Kolmogorov equations for the partial characteristic functions in the steady state regime:

$$\left\{ \begin{array}{l} -\lambda H_0(u_1, u_2) + \lambda e^{2ju_1} H_1(u_1, u_2) + H_2(u_1, u_2)[\mu_2 + \lambda e^{ju_1} e^{ju_2}] + \\ + j\sigma_1 \frac{\partial H_0(u_1, u_2)}{\partial u_1} + j\sigma_2 \frac{\partial H_0(u_1, u_2)}{\partial u_2} - \\ - j\sigma_1 e^{ju_1} \frac{\partial H_1(u_1, u_2)}{\partial u_1} - j\sigma_2 e^{ju_1} \frac{\partial H_1(u_1, u_2)}{\partial u_2} - \\ - j\sigma_2 e^{ju_2} \frac{\partial H_2(u_1, u_2)}{\partial u_2} - j\sigma_1 e^{ju_2} \frac{\partial H_2(u_1, u_2)}{\partial u_1} = 0, \\ \lambda H_0(u_1, u_2) - (\lambda + \mu_1) H_1(u_1, u_2) - j\sigma_1 e^{-ju_1} \frac{\partial H_0(u_1, u_2)}{\partial u_1} + \\ + j\sigma_1 \frac{\partial H_1(u_1, u_2)}{\partial u_1} + j\sigma_2 \frac{\partial H_1(u_1, u_2)}{\partial u_2} = 0, \\ -(\lambda + \mu_2) H_2(u_1, u_2) - j\sigma_2 e^{-ju_2} \frac{\partial H_0(u_1, u_2)}{\partial u_2} + \\ + j\sigma_2 \frac{\partial H_2(u_1, u_2)}{\partial u_2} + j\sigma_1 \frac{\partial H_2(u_1, u_2)}{\partial u_1} + \mu_1 H_1(u_1, u_2) = 0. \end{array} \right. \quad (1)$$

4. Asymptotic analysis under the long delay condition

Denote

$$\sigma_k = \gamma S_k.$$

System (1) will be solved by applying the method of asymptotic analysis under the long delay condition $\gamma \rightarrow 0$.

4.1. Asymptotic of the first order

Substituting

$$\sigma_k = \varepsilon S_k, u_k = \varepsilon w_k, H_k(u_1, u_2) = F_k(w_1, w_2, \varepsilon)$$

into the system (1) we obtain:

$$\left\{ \begin{aligned} & -\lambda F_0(w_1, w_2, \varepsilon) + \lambda e^{2j\varepsilon w_1} F_1(w_1, w_2, \varepsilon) + jS_1 \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_1} + \\ & + (\mu_2 + \lambda e^{j\varepsilon w_1} e^{j\varepsilon w_2}) F_2(w_1, w_2, \varepsilon) + jS_2 \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_2} - \\ & - jS_1 e^{j\varepsilon w_1} \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_1} - jS_2 e^{j\varepsilon w_1} \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_2} - \\ & - jS_2 e^{j\varepsilon w_2} \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_2} - jS_1 e^{j\varepsilon w_2} \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_1} = 0, \\ & \lambda F_0(w_1, w_2, \varepsilon) - (\lambda + \mu_1) F_1(w_1, w_2, \varepsilon) + jS_1 \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_1} - \\ & - jS_1 e^{-j\varepsilon w_1} \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_1} + jS_2 \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_2} = 0, \\ & -(\lambda + \mu_2) F_2(w_1, w_2, \varepsilon) - jS_2 e^{-j\varepsilon w_2} \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_2} + \\ & + jS_2 \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_2} + jS_1 \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_1} + \\ & + \mu_1 F_1(w_1, w_2, \varepsilon) = 0. \end{aligned} \right. \quad (2)$$

Theorem 1 *The limit value $F_k(w_1, w_2)$ of the solution $F_k(w_1, w_2, \varepsilon)$ of the system (2) has the form:*

$$F_k(w_1, w_2) = R_k \exp \{ jw_1 a_1 + jw_2 a_2 \},$$

where the parameters a_1, a_2 and R_k are defined as follows:

$$R_1 = \frac{\lambda}{\mu_1}, R_2 = \frac{\lambda}{\mu_2}, R_0 = 1 - \frac{\lambda}{\mu_1} - \frac{\lambda}{\mu_2},$$

$$a_1 = \frac{\lambda(2R_1 + R_2)(R_0 - R_2) + R_2 R_1}{S_1(R_0 - R_1)(R_0 - R_2) - R_2 R_1},$$

$$a_2 = \frac{\lambda R_2(2R_1 + R_2) + (R_0 - R_1)}{S_2(R_0 - R_1)(R_0 - R_2) - R_2 R_1}.$$

Parameters R_1, R_2, R_0 represent the stationary probabilities of the states of the device, a_1, a_2 are the average numbers of requests in the first and the second blocks of the orbit correspondingly.

Theorem 2 *The throughput value of M/GI/1 retrial system with conflicts and afterservice is defined by*

$$S = \frac{1}{2}.$$

4.2. Asymptotic of the second order

In the system (1) let us denote

$$H_k(u_1, u_2) = H_k^{(2)}(u_1, u_2) \exp \left\{ j u_1 \frac{a_1}{\gamma} + j u_2 \frac{a_2}{\gamma} \right\}.$$

Substituting

$$\gamma = \varepsilon^2, \sigma_k = \varepsilon^2 S_k, u_k = \varepsilon w_k, H_k^{(2)}(u_1, u_2) = F_k(w_1, w_2, \varepsilon)$$

into the system (1) we get:

$$\left\{ \begin{array}{l} -\lambda F_0(w_1, w_2, \varepsilon) + \lambda e^{2j\varepsilon w_1} F_1(w_1, w_2, \varepsilon) + j\varepsilon S_1 \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_1} + \\ + (\mu_2 + \lambda e^{j\varepsilon w_1} e^{j\varepsilon w_2}) F_2(w_1, w_2, \varepsilon) + j\varepsilon S_2 \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_2} - \\ - j\varepsilon S_1 e^{j\varepsilon w_1} \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_1} - j\varepsilon S_2 e^{j\varepsilon w_1} \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_2} - \\ - j\varepsilon S_2 e^{j\varepsilon w_2} \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_2} - j\varepsilon S_1 e^{j\varepsilon w_2} \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_1} - \\ - S_1 a_1 F_0(w_1, w_2, \varepsilon) - S_2 a_2 F_0(w_1, w_2, \varepsilon) + \\ + S_1 a_1 e^{j\varepsilon w_1} F_1(w_1, w_2, \varepsilon) + S_2 a_2 e^{j\varepsilon w_1} F_1(w_1, w_2, \varepsilon) + \\ + S_2 a_2 e^{j\varepsilon w_2} F_2(w_1, w_2, \varepsilon) + S_1 a_1 e^{j\varepsilon w_2} F_2(w_1, w_2, \varepsilon) = 0, \\ \lambda F_0(w_1, w_2, \varepsilon) - (\lambda + \mu_1) F_1(w_1, w_2, \varepsilon) + j\varepsilon S_1 \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_1} - \\ - j\varepsilon S_1 e^{-j\varepsilon w_1} \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_1} + S_1 a_1 e^{-j\varepsilon w_1} F_0(w_1, w_2, \varepsilon) + \\ + j\varepsilon S_2 \frac{\partial F_1(w_1, w_2, \varepsilon)}{\partial w_2} - S_1 a_1 F_1(w_1, w_2, \varepsilon) - \\ - S_2 a_2 F_1(w_1, w_2, \varepsilon) = 0, \\ -(\lambda + \mu_2) F_2(w_1, w_2, \varepsilon) - j\varepsilon S_2 e^{-j\varepsilon w_2} \frac{\partial F_0(w_1, w_2, \varepsilon)}{\partial w_2} + \\ + j\varepsilon S_2 \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_2} + j\varepsilon S_1 \frac{\partial F_2(w_1, w_2, \varepsilon)}{\partial w_1} + \\ + \mu_1 F_1(w_1, w_2, \varepsilon) + S_2 a_2 e^{-j\varepsilon w_2} F_0(w_1, w_2, \varepsilon) - \\ - S_1 a_1 F_2(w_1, w_2, \varepsilon) - S_2 a_2 F_2(w_1, w_2, \varepsilon) = 0. \end{array} \right.$$

Theorem 3 *The limit value $F_k(w_1, w_2)$ of the solution $F_k(w_1, w_2, \varepsilon)$ of the system (12) has the form:*

$$F_k(w_1, w_2) = R_k \exp \left\{ \frac{(jw_1)^2}{2} \kappa_{11} + jw_1 jw_2 \kappa_{12} + \frac{(jw_2)^2}{2} \kappa_{22} \right\},$$

where parameters R_k are defined by expression 7 and parameters κ_{ii} are the solution of the following system:

$$\left\{ \begin{aligned} & \frac{B}{\mu_2} \left\{ S_1 a_1 - \lambda - \frac{(2\lambda + S_2 a_2)(\lambda + S_1 a_1)}{\mu_1 + S_2 a_2} \right\} + A \frac{2\lambda + S_2 a_2}{\mu_1 + S_2 a_2} - \\ & - R_1 (2\lambda + \frac{1}{2})(S_1 a_1 + S_2 a_2) + S_1 \kappa_{11} (R_0 - R_1) - \\ & - \frac{1}{2} \lambda R_2 - S_2 R_1 \kappa_{12} = 0, \\ & \frac{C}{\mu_2} \left\{ S_2 a_2 \frac{\lambda + S_1 a_1}{\mu_1 + S_2 a_2} - (\lambda + S_1 a_1) \right\} - D \frac{S_2 a_2}{\mu_1 + S_2 a_2} - \frac{1}{2} S_2 a_2 R_0 - \\ & - \frac{1}{2} R_2 (\lambda + S_1 a_1 + S_2 a_2) + S_2 (R_0 - R_2) \kappa_{22} - S_1 R_2 \kappa_{12} = 0, \\ & \frac{B}{\mu_2} \left\{ -S_1 a_1 - \lambda + \frac{S_2 a_2 (\lambda + S_1 a_1)}{\mu_1 + S_2 a_2} \right\} - A \frac{(\lambda + S_1 a_1) S_2 a_2}{\mu_1 + S_2 a_2} + \\ & + \frac{C}{\mu_2} \left\{ S_1 a_1 - (2\lambda + S_2 a_2) \frac{\lambda + S_1 a_1}{\mu_1 + S_2 a_2} - \lambda \right\} - \lambda R_2 + \\ & + D \frac{(\lambda + S_1 a_1)(2\lambda + S_2 a_2)}{\mu_1 + S_2 a_2} + [(S_1 + S_2) R_0 - S_1 R_1 - S_2 R_2] \kappa_{12} - \\ & - S_2 R_1 \kappa_{22} - S_1 R_2 \kappa_{11} = 0. \end{aligned} \right.$$

where A, B, C, D are defined as follows:

$$\begin{aligned} A &= S_1 a_1 R_0 + S_1 (R_1 - R_0) \kappa_{11} + S_2 R_1 \kappa_{12}, \\ B &= -R_1 (2\lambda + S_1 a_1 + S_2 a_2) - \lambda R_2 + S_1 (R_0 - R_1 - R_2) \kappa_{11} + S_2 (R_0 - R_1 - R_2) \kappa_{12}, \\ C &= -R_2 (\lambda + S_1 a_1 + S_2 a_2) + S_1 (R_0 - R_1 - R_2) \kappa_{12} + S_2 (R_0 - R_1 - R_2) \kappa_{22}, \\ D &= S_1 (R_1 - R_0) \kappa_{12} + S_2 R_1 \kappa_{22}. \end{aligned}$$

5. Conclusions

Thus, in this paper we considered M/GI/1 retrial system with conflicts and afterservice. We found the throughput value, the stationary probabilities of the states of the device, the average number of requests in the blocks of the orbit and the characteristic function of the number of requests in the blocks of the orbit for this system.

Acknowledgments

The reported study was funded by RFBR according to the research project No. 16-31-00292 mol-a.

References

1. *Falin G. I., Sukharev Yu. I.* On single-line queue with double connection – All-Union Institute for scientific and Technical Information. Moscow, 1985.
2. *Choi B. D., Shin Y. W., Ahn W. C.* Retrial Queues with collision arising from unslotted CSMA/CD protocol – Queueing Systems. – 1992 – no. 11. – P. 335–356.
3. *Krishna Kumar B., Vijayalakshmi G., Krishnamoorthy A., Sadiq Basha S.* A single server feedback retrial queue with collisions. – Computer and Operations Research. – 2010. – no. 37. – P. 1247–1255.
4. *Artalejo J. R., Gomez-Corral A.* Retrial Queueing Systems. – Springer. – 2008.
5. *Jeongsim Kim.* Retrial queueing systems with collision and impatience. – Commun. Korean Math. Soc. –25(2010). –no. 4. –P. 647-653

УДК 004.932

Повышение эффективности балансировки трафика в сети на основе оценки концентрации внимания пользователя

Д.В. Окунева[†], К.С Прошутинский*

* *Кафедра сетей связи и передачи данных, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, ул. пр. Большеви-ков, д.22, Санкт-Петербург, Россия, 193232*

[†] *Кафедра программной инженерии и вычислительной техники, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, ул. пр. Большеви-ков, д.22, Санкт-Петербург, Россия, 193232*

Аннотация. В настоящее время наблюдается рост инфокоммуникационных услуг, в основе которых лежит передача видео. Этот тип трафика предъявляет высокие требования к показателям качества обслуживания. В то же время пользователи, высказывают повышенные требования к качеству восприятия видеослуж. Для предоставления видеослуж с надлежащим качеством оператору необходимо пересмотреть подходы к управлению трафиком, распределению сетевых ресурсов, балансировке нагрузки. В статье предлагается способ управления передачей видеопотоков в зависимости от концентрации внимания пользователя. Видеопоток транслируется с хорошим качеством, в случае заинтересованности пользователя в просмотре, или с низким качеством, в случае её отсутствия, что позволяет экономить пропускную способность сети на данном участке сети.

Ключевые слова: качество обслуживания, качество восприятия, управление трафиком, видеослуж, до-полненная реальность, концентрация внимания.

1. Введение

Контроль и управление потоками трафика является одной из сложных задач управления в IP-сетях. Сегодня эта задача становится еще более актуальной, поскольку число пользователей видеослуж растет [3] и в то же время отмечается рост требований к качеству передачи видео [1,2]. Следует отметить, что пользовательские терминалы могут воспроизводить видео в различных форматах с разным разрешением, это вынуждает операторов передавать по сети потоки одного и того же видео в разном качестве. Это приводит к увеличению объема видеотрафика в сетях связи, и, как следствие, к необходимости пересмотра механизмов управления сетями. При этом серьезной проблемой, в случае существенного роста объемов видеотрафика, может стать ограниченность физического ресурса сети. До сих пор эта проблема решалась

за счет увеличения числа каналов и их пропускной способности, обеспечиваемой изменением среды передачи и переходом на новые протоколы взаимодействия, оптимизации маршрутов, классификации трафика и управления потоками [7, 8]. В тоже время невозможно бесконечно увеличивать пропускную способность каналов. Для того чтобы удовлетворить запросы пользователей и не строить новую сетевую инфраструктуру, необходимо совершенствовать механизмы управления потоком и балансировки нагрузки. Для решения поставленной задачи предлагается метод управления интенсивностью видеопотока в сети на основе анализа концентрации внимания потребителя услуги. Эти данные оператор может использовать не только для управления трансляцией видеопотоков, но и для определения заинтересованности пользователя предлагаемым контентом, что позволит в дальнейшем оценить качество восприятия видеослужбы [4]. В сетях проблема распределения ресурсов решается постоянно, как на этапе планирования и проектирования сетей связи, так и при эксплуатации. Сложности возникают при появлении в сети избыточной нагрузки, связанной с появлением новых видов услуг [5]. В результате, сформировавшийся запас ресурса сети быстро иссекает и образовывается дефицит, оказывающий влияние на качество восприятия (Quality of Experience – QoE) [9, 10]. В тоже время, известно, что видеотрафик, предъявляющий высокие требования к значениям пропускной способности и другим показателям QoE, не всегда необходим пользователю в наилучшем качестве. Если пользователь включил телевизор как «фон» и не смотрит его, можно говорить о допустимом снижении качества передаваемого видеопотока за счет переключения на поток с более низкой скоростью передачи и низким качеством видео. В результате пользователь не замечает разницы, а оператор экономит транспортный ресурс сети и может использовать его для передачи трафика другого пользователя. Целью исследования является изучение подходов повышения эффективности управления транспортным ресурсом сети на основе распознавания концентрации внимания пользователя, что позволит, не снижая качества восприятия услуги, минимизировать объем видеотрафика в сети.

2. Структура стенда

Для решения поставленной задачи была собрана лабораторная установка. Структура лабораторного стенда, использованного для проведения эксперимента по балансировке трафика в телекоммуникационных сетях, представлена на рис.1.

Лабораторный стенд, моделирующий телекоммуникационную сеть, представлен локальной компьютерной сетью, включающей клиента (модель пользователя) и сервер (модель провайдера). Модель пользователя включает в себя терминал и подключенную к нему камеру «дополненной реальности» Intel RealSense 3D [6], которая распознает эмоции, движения, жесты пользователя. Модель провайдера представляет собой компьютер, который выполняет роль сервера и генерирует

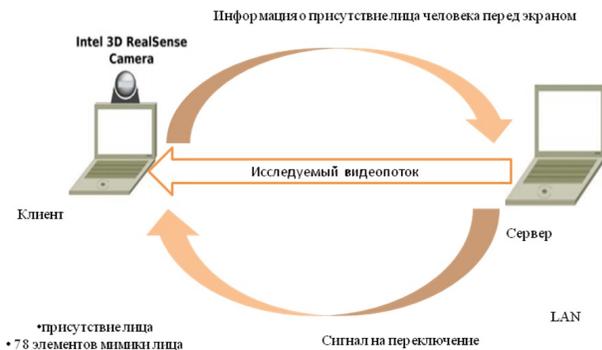


Рис. 1. Схема экспериментальной установки балансировки трафика

в сеть видеотрафик. Разработанное в ходе проведения эксперимента программное обеспечение имеет клиентсерверную архитектуру. Задачей клиентской части является получение информации о наличии или отсутствии лица перед экраном и передачи этих данных для анализа на сервер. Сбор информации осуществляет камера Intel RealSense 3D. Она обладает лучшим в своем классе датчиком глубины сцены и сенсором с разрешением 1080p, способна сканировать предметы и получать их трехмерное изображение, различать передний и задний планы, распознавать сложные жесты вплоть до движения пальцев, мимику и черты лица для определения эмоций. В описываемом эксперименте были задействованы следующие возможности камеры: -определение присутствия лица человека в диапазоне работы видеокамеры; -распознавание 78 точек лица для определения концентрации внимания.

3. Описание эксперимента

С сервера одновременно запускаются видеопотоки высокого и низкого качества. Начинается процесс трансляции на стороне пользователя. Одновременно с этим включается камера, которая определяет лицо человека и его положение относительно экрана. Если лицо обнаружено, и определено, что оно смотрит на экран, то видеопоток передается с высоким качеством. В случае, если камера фиксирует поворот головы и не может определить лицо перед экраном, программное обеспечение клиента, обработав эту информацию, отправляет сообщение на сервер. Сервер анализирует полученные данные и формирует команду на переключение видеопотока с высоким качеством изображения на видеопоток с низким качеством. После того как камера вновь обнаруживает лицо пользователя, и с клиента приходит соответствующее



а)



б)

Рис. 2. Пропускная способность видеопотоков высокого и низкого качества: а) для потоков 1080p и 180p; б) для потоков 1080p и 720p

сообщение, программное обеспечение сервера выполняет обратное переключение. Этот процесс реализуется в течение всего сеанса передачи видеoinформации. Результаты проведенного исследования отражены в графиках, представленных на рисунке 2а и 2б.

Графики демонстрируют принципиальную возможность экономии пропускной способности сети за счет переключения видеопотоков с высоким качеством на низкое. Тем самым за счет балансировки нагрузки достигается эффективное использование транспортного ресурса сети.



Рис. 3. Схема экспериментальной установки балансировки трафика, при наличии сторонней нагрузки в сети

При переключении видеопотока с высокого разрешения 1080p на видеопоток с низким разрешением происходит экономия пропускной способности канала от максимального 14 Мбит/с до минимального значения в 4 Мбит/с. Эта экономия пропускной способности канала может быть использована для удовлетворения запросов других пользователей сети и передачи трафика этих пользователей с более высоким качеством обслуживания. Описанный выше эксперимент моделирует работу сети без сторонней нагрузки, когда в сети присутствует только видеотрафик генерируемый сервером.

В реальной ситуации нагрузка в сети определяется суммарным объемом трафика, исходящим от провайдера к множеству пользователей. Профили пользователей одного провайдера отличаются в зависимости от типов услуг и времени их запроса. А так как каждый вид трафика характеризуется различной скоростью передачи, то он предъявляет различные требования к характеристикам сети, что создает дополнительные трудности для решения проблемы эффективной балансировки нагрузки.

Для моделирования данной ситуации, формируется агрегированный поток включающий в себя несколько видеопотоков с различными характеристиками. В реальности такая ситуация возможна в любом многоквартирном доме, при предоставлении «Triple Play» услуг. Программное обеспечение IxChariot, использованное в данном эксперименте, позволяет имитировать нагрузку канала суммарным трафиком,



а)



б)

Рис. 4. Пропускная способность видеопотоков высокого и низкого качества: а) для потоков 1080p и 180p; б) для потоков 1080p и 720p

потребляемым пользователями одной сети. Схема экспериментальной установки была видоизменена и представлена на рис.3. В процессе эксперимента имитировался трафик, поступающий от двухсот абонентов. Результаты проведенного исследования отражены в графиках, представленных на рисунке 4а и 4б.

Экономии пропускной способности канала можно отметить и при работе сети под нагрузкой. Например, как видно на графике 4а, экономия может достигать 56 Мбит/с. В результате исследования была разработана программная система, позволяющая минимизировать объём видеотрафика в сети, не снижая качества восприятия услуги. Данная цель была достигнута, используя метод управления потоком на основе распознавания концентрации внимания и подтверждена результатами тестирования представленными на рисунках 3 и 4.

4. Заключение

В статье предложен метод управления потоком на основе оценки концентрации внимания. Проведен ряд экспериментов на модельной сети, в ходе которых, установлена эффективность использования предложенного метода, что позволит в дальнейшем рационально распределять ресурсы сети и, как следствие, повысить QoE пользователей. Дальнейшее развитие может состоять в реализации полной системы интерактивного IPTV и разработке новых методов, позволяющих оценивать не только присутствие зрителя, но и состояние активности пользователя: спит, читает книгу (не смотрит на экран терминала), а также степень интереса пользователя к той или иной телепрограмме.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00209 мол а "Разработка принципов интеграции технологии дополненной реальности и Интернета Вещей."

Литература

1. Recommendation G.1011. Reference Guide to Quality of Experience Assessment Methodologies. ITU-T, Geneva, May 2013.
2. *Kirichek R. V.* Prospects for the development of video services with recognition /R.V emotions. R V Kirichek, M. A. Makolkina, D. A. Shangina, A. A. Sholokhov. - Moscow: Publishing House of the Information Technology and Telecommunications, 2015. - 139p.
3. *Koucheryavy A. E.* public communication network. Tendencies of development and methods of calculation / A. E. Koucheryavy, A. I. Paramonov, . - M.: Publishing house ZNIIS, 2008. - 296 p.
4. *Makolkina M.A.* Evaluation of the quality video experience based on emotion recognition // Telecommunications. - 2015. - № 9. - S. 24-28. - ISSN: 0013-5771 / HAC, RISC.
5. *Koucheryavy A. E.* Internet of Things / A. E. Koucheryavy Telecommunications. - 2013. - No 1. - S. 21-24.
6. *Alekseeva D. D. , M. A. Makolkina* augmented reality technology // Information Technology and Telecommunications. - 2015. -

- № 2 (10). - S. 101-110. - ISSN: 2307-1303 / RISC. - URL: <http://www.sut.ru/doci/nauka/review/2-15.pdf>.
7. *Makolkina M. A.* Analysis of subjective methods for evaluating perceptual quality in ip-networks // 70 Anniversary All-Russia Scientific Conference, dedicated to the radio - SPb. : SPbGEU "LETI"na. IN AND. Ulyanov (Lenin), 2015. - P. 196-198 / RISC.
 8. *Makolkina M.A., Saint Jean V. Kulik V.A.* Network Model to assess the quality perception of services // Information Technology and Telecommunications. - 2015. - № 3 (11). - S. 40-44. - ISSN: 2307-1303 / RISC. - URL: <http://www.sut.ru/doci/nauka/review/3-15.pdf>.
 9. *Kirichek R., Koucheryavy A.* Internet of things laboratory test bed // International conference on wireless communication, networking and application. wena 2014 — Heidelberg : Springer, 2014. — C. / SCOPUS.
 10. *Makolkina M., Prokopyev A., Paramonov A., Koucheryavy A.* (2014) The quality of experience subjective estimations and the hurst parameters values interdependence. IN Balandin S., Andreev S., Koucheryavy Y. (Eds.), Internet of things, smart spaces, and next generation networks and systems.

UDC 004.932

Improving of the traffic balancing efficiency on the base of estimations of user attention concentration

D. V. Okuneva[†], K. S. Proshutinskiy*

* *Department of Infocommunication Systems , The Bonch-Bruевич Saint-Petersburg State University of Telecommunications, str. pr. Bolsheviks, d.22, St. Petersburg, Russia, 193232*

[†] *Department of Programming Engineering and Computer Facilities, The Bonch-Bruевич Saint-Petersburg State University of Telecommunications, str. pr. Bolsheviks, d.22, St. Petersburg, Russia, 193232*

There is currently a growing a communication services, which are based on the transfer of the video. This type of traffic makes high quality of service performance requirements, and users, in turn, have high requirements for video services. In order to provide good quality video services with the operator need to review approaches to traffic management, allocation of network resources, load balancing. The article describes the approach based on the concentration of attention of the user, broadcast video stream with good quality, in the case of interest to the user in viewing, or poor quality, in case of its absence, thus saving bandwidth on a given network area.

Keywords: QoS, QoE, traffic control, video services, augmented reality, concentration of attention.

UDC 004.4

Simulation of Medical Sensor Nanonetwork Applications Traffic

R. Pirmagomedov, I. Hudoev, D. Shangina

*Department of Telecommunication Networks and Data Transmission
Saint-Petersburgs State University of Telecommunications
Bolshevikov pr. 22/1, Saint-Petersburg, 193232, Russia*

Abstract. The article deals with simulation of traffic from sensor nanonetwork applications. The paper presents results of nanonetwork applications traffic simulation. Simulation based on the traffic models developed for M2M. In the simulation considered the possibility of gateway working in two modes: without processing messages received from nanonetwork and with it. Typical architecture of nanonetwork medical applications involving the use of remote Internet servers, was described. The results of traffic flow simulation were analysed on the self-similarity properties.

Keywords: nanonetworks, internet of things, traffic modeling.

1. Introduction

The progress at Internet of Nano-things [1, 2], supposes development of many applications, using nanonetwork structures in different spheres of human life whether it be industrial sphere, military sphere or everyday human life [3]. Medicine is one of the promising fields of nanonetwork structures use [4]. It is expected that nanonetwork applications will supplement the Internet of Things technologies, which have already existed in medical sphere [5, 6] and will open new opportunities for diagnosing and diseases treating [7, 8], environmental monitoring, performing of surgical operation for tissue reformation at molecular and DNA levels [9], making of smart medicines [10, 11] and so on. Development of such applications is supposed in three contexts: inside of human body, on the surface of human body and in human inhabitations (apartments, car, hospital room). It is planned that all above-listed applications will have the opportunity to transfer and receive the information from remote servers using the Internet resources [12]. Today nanonetwork structures can't make full convergence with traditional communication networks and now exist in the form of autonomous structures which were created to solve the narrow-purpose problems [10, 11]. However, active researches carried out at this theme, led to the conclusion that in the future the question of nano- and traditional networks union will be decided [8]. The development and putting in operation of many nanoapplications, generating additional traffic flows in the Internet, will require reconsideration of established paradigms in the sphere of telecommunication. To forecast necessary changes of traditional networks, we need to research the behavior of nanonetwork applications traffic flows. First of all, it is necessary to answer the following questions:

1. What type of traffic flows will be generated by these nanonetwork applications.

2. What characteristics these traffic flows will have.

In view of the fact that the Internet of Nano-things is a logical continuation of the Internet of Things conception [12], it would be logically to suppose that characteristics of traffic generated by nanonetwork applications will be quite similar with the Internet of Things traffic. According to it, patterns have already been suggested and investigated in the context of the Internet of Things [13–15] may be also applied for Internet of Nano-things. Communication of nanonetwork structures with computer center (remote server) is suitable for the context of M2M conception [16,17], being its new manifestation. In contradistinction from the Internet of Things traffic patterns, in nanonetwork applications sensor readouts are sent to server not immediately, but firstly get through the nanonetwork, which certainly has an influence on the transfer characteristics. One more factor having influence on traffic parameters is availability of intermediated node (gateway) between nanonetwork and the Internet. Gateway working algorithms, its opportunities of changing traffic parameters and message preprocessing should be taken into account during development and investigation of nanonetwork applications traffic patterns.

In this work we develop the pattern for nanosensor network with its connection with the Internet and make a comparison with the pattern of monitoring system and supervisory control traffic flow. The article has the following structure. In section 2 different aspects of nanonetwork applications traffic simulation depending on their functionality and field of application are considered. In the third part peculiarities of traffic simulation for sensor nanonetwork are described. In the fourth part traffic pattern for nanosensor network is developing. In the fifth section, with the help of simulation modelling system the investigation of this pattern and comparison of results with the pattern of monitoring system and supervisory control traffic flow are described.

2. Architectures of Nanonetwork Medical Applications

Nanonetwork medical applications should solve a wide range of problems from sensor readouts capture, situated as well as in environment, and directly inside the human body, to tissue reformation and microsurgical operations. Depending on solving problems nanonetwork applications will have different functionality and therefore traffic flows of such applications will also be different. According to their functional peculiarities nanonetwork applications can be divided into a few basic classes:

- 1) Autonomous application. It means autonomous work of nanonetwork developed in the environment, inside the human body or on the surface of human body. Data interchange with remote server is not supposed for nanoapplications of this class or data interchange will be restricted (technical message interchange in cases of critical condition, about workability of the network and so on). So, traffic between nanonetwork and

remote server either is absent at all or relatively small volume of traffic is generated because application responds to the critical event. The events, caused generation of traffic are relatively rare.

2) Sensor nanonetwork. It is the application based on nanosensor network use, in which nanosensors are activated on demand of server or local computer center. Such applications are used for periodical capture of information about the state of human health or about the state of environment. Traffic of such systems will represent determinated packets flow consists of server queries and nanonetwork application answers.

3) Sensor-actuator application. It is nanonetwork application which aims not only for capture of the information about state of organism with the help of nanosensors, but also at influence on processes in body with the help of nanoactuators (for example, remote microscopic surgical operations in human body). Applications of this type are very sensitive to delay and signal distortions and can't work off-line. Traffic of such systems can't be classified as M2M traffic, because of the presence of the operator (experienced surgeon), who performs the operation.

Final application can have mixed functionality and characteristics of different classes. Taking into account that sensor-actuator applications are very demanding to QoS parameters, such systems should be developed locally, because it is very difficult to provide necessary QoS parameters using connection via the Internet. In view of this fact we can consider traffic of such applications only at local area networks. Hereinafter we will consider only sensor nanonetworks traffic.

3. The Peculiarities of Sensor Nanonetwork Traffic Simulation

The work of sensor nanonetwork builds on the master-slave principle (see fig. 1), while data capture realized according to predetermined rules. Traffic generated by this application will be pseudodetermined.

There is probability of message loss in nanonetwork applications - p because of realization difficulties at physical level and imperfection of routing algorithms. "Normal" probability of message loss for some nanonetwork technologies can account for 95 percents. Message loss can happen as well as during scanning of sensors by gateway, and during data transfer from sensors to gateway after receiving of query. Therefore, integrated probability of sensor readouts transfer to gateway P_{data} will be expressed by (1).

$$P_{data} = (1 - p)(1 - p) = 1 - 2p + p^2 \quad (1)$$

p - is probability of message loss in nanonetwork.

In spite of probability of message loss, there are different working algorithms of gateway during scanning of sensors and during their data transfer to remote server:

1. Without processing of messages from sensors

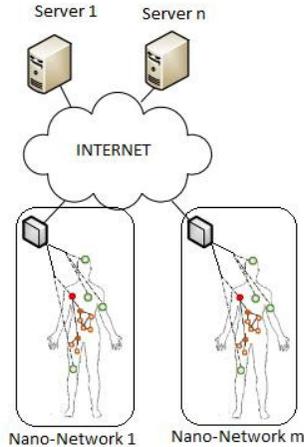


Figure 1. General structure of sensor nanonetwork application

2. With processing of messages from sensors

Working mode without processing of messages from sensors supposes all the messages transferred from sensor to be sent to the remote server directly. In this mode, sensors readouts are not processed and transferred to remote server from each sensor separately. Gateway, working in this mode, has high response speed concerning the second working mode and small buffer capacity. High traffic on communication channel between gateway and remote server can be referred to the disadvantages of such working mode. Traffic in this case is characterized by a great amount of packets containing readouts from only one sensor.

The mode with processing of messages from sensors supposes storage and preprocessing of sensors readouts before send them to remote server. In this mode efficiency of use of network resources increases during data interchange between gateway and server at the expense of data part in packets becomes considerably bigger, moreover frequency and amount of messages decreases. However, delay from the moment of server query arrival to sending of data from gateway increases in comparison with the first mode, basically because of additional time which requires for storage and processing of data from sensors. The architecture of gateway in this case becomes more difficult due to high system requirements.

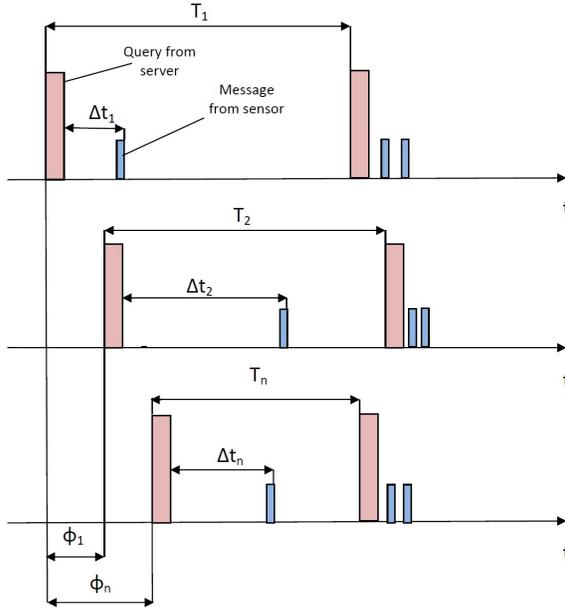


Figure 3. Sensor network traffic pattern for gateway mode with processing of messages from sensors

queries and nanonetwork application answers (fig. 3). Time parameters of system flow, in general, can be defined by time-table and have definite repeat period T_i .

Time from the moment of server query arrival to sending readouts to server we shall designate Δt . This time gateway inquires sensors located in nanonetwork and processes received information. Sensors readouts are sent to server. The amount of messages sent by gateway to server will depend on the volume of information received from sensors. Limited quantity of information can be transmitted in one message D_{info} . Quantity of useful information can be calculated by (2).

$$D_{info} = V_{max} - D_{service} \quad (2)$$

where V_{max} is maximal packet size for this application, $D_{service}$ is a volume of ordering information. In our work we will suppose that $V_{max} = 1500$ bytes (Ethernet frame size), $D_{service} = 62$ bytes (the sum of headers sizes Ethernet, IP, TCP). Using (2), 1438 bytes of useful information can be transferred. The volume of information, received from one sensor we will

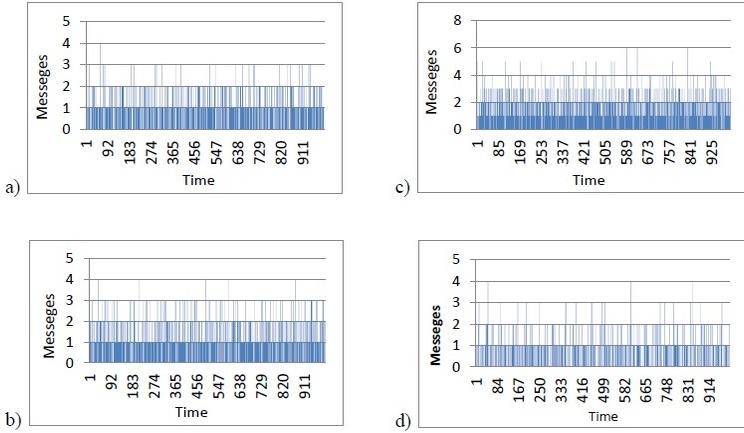


Figure 4. Traffic flow for gateway mode with preprocessing of messages a) $\Delta t_i = 0,2 * T_i$ b) $\Delta t_i = 0,4 * T_i$ c) $\Delta t_i = 0,6 * T_i$ d) $\Delta t_i = 0,8 * T_i$

consider as $v_0 = 16$ bytes (4 bytes for sensors readouts + 4 bytes for sensor address in this nanonetwork + 4 bytes for sensor type). Using above-listed considerations we can transfer readouts of 89 sensors in one packet. If it is necessary to transfer readouts of greater amount of sensors, we shall send one more message. In general case, general amount of packets which would be necessary to transfer readouts of all the sensors N_{pages} is calculated by formula (3)

$$N_{pages} = \text{ceiling}(N_{v0}/D_{info}) \quad (3)$$

where N is the amount of sensors in an application. (ceiling is rounding upward), v_0 is the data volume of one sensor.

5. Results of Simulation

If there are n nanonetwork applications, so for each working mode of gateway, at unchanged phase shifts between moments of queries (answers) arrival φ_i for $i = 1..n$ general traffic will also represent determined periodical process with a period equal to least common multiple of all scanning periods. Also at the beginning of each period of time T_i , from i server to j application a query for receiving sensors readouts arrives. Δt_i coefficient for each of systems is selected equal to $0,2 * T_i$; $0,4 * T_i$; $0,6 * T_i$ and $0,8 * T_i$. The amount of sensors located in one nanonetwork application is $N = 100$. For application pattern with gateway without preprocessing of messages

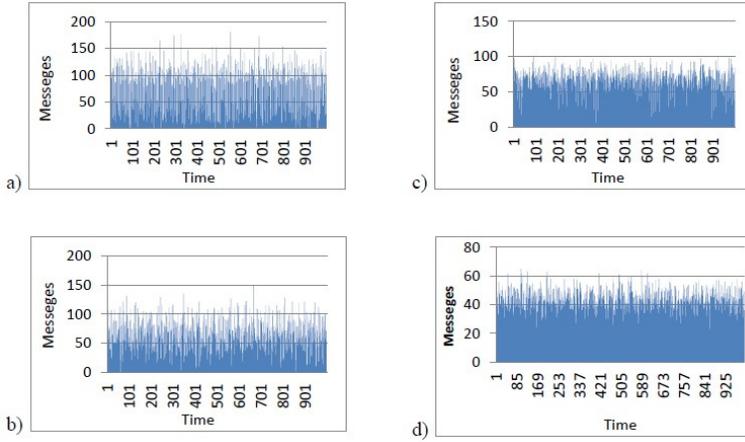


Figure 5. Traffic flow for gateway mode without preprocessing of messages a) $\Delta t_i = 0,2 * T_i$ b) $\Delta t_i = 0,4 * T_i$ c) $\Delta t_i = 0,6 * T_i$ d) $\Delta t_i = 0,8 * T_i$

selection of time for sending readouts of one sensor to server is drawing at random in the limits from 0 to Δt_i (selection of time for sending in a pattern is given by uniform law). Results of simulation are given at the figures 4 and 5.

Flows received on the results of simulation were checked for self-similarity features. For gateway mode with preprocessing of messages from sensors Hurst coefficient $H=0.12$ at $\Delta t_i = 0.2 * T_i$, $H=0.11$ at $\Delta t_i = 0.4 * T_i$, $H=0.19$ at $\Delta t_i = 0.6 * T_i$, $H=0.19$ at $\Delta t_i = 0.8 * T_i$. For gateway mode without preprocessing of messages from sensors Hurst coefficient made up $H=0.21$ at $\Delta t_i = 0.2 * T_i$, $H=0.24$ at $\Delta t_i = 0.4 * T_i$, $H=0.18$ at $\Delta t_i = 0.6 * T_i$, $H=0.27$ at $\Delta t_i = 0.8 * T_i$. From received results we can see that in all cases Hurst coefficient is considerably less than 0.5, therefore traffic in each experiment is self-similar stochastic process, corresponding to the class of antipersistent processes. This result was received for all calculated data series for different correlations of periods between T queries and time of nanonetwork application reaction Δt_i .

6. Conclusions

In this work the analysis of self-similarity feature of traffic generated by nanonetwork applications was carried out. It is known that self-similarity feature of traffic influence on QoS in network [18] (for many types of nanonetwork applications, guaranteed short time of packets delivery or/and level of packets loss are critical). Taking into account that

there are two working mode of gateway for which traffic has self-similarity features, we proved possibility of using gateway working mode without considerable deterioration of QoS. It is can be useful for nanonetworks real-time application development.

References

1. *Bari N., Berkovich S., Ganapathy M.* Internet of Things as a Methodological Concept, Computing for Geospatial Research and Application (COM.Geo), 2013 Fourth International Conference, p.48-55, 2013.
2. *Akyildiz A. F., Brunetti F., Blazquez C.* A New Communication Paradigm. — Computer Networks (Elsevier) Journal, Vol. 52, p.2260-2279, August, 2008.
3. *Shyamkumar P., Rai P., Oh S., Ramasamy M., Harbaugh R., Varadan V.* Wearable Wireless Cardiovascular Monitoring Using Textile-Based Nanosensor and Nanomaterial Systems, Electronics, vol. 3, no. 3, p.504-520, 2014.
4. *Kirichek R., Pirmagomedov R., Glushakov R., Koucheryavy A.* Live Substance in Cyberspace - Biodriver System, Proceedings, 18th International Conference on Advanced Communication Technology (ICACT) 2016. — Phoenix Park, Korea, p.274-278, 2016.
5. *Seyedi M., Kibret B., Lai D. and Faulkner M.* A Survey on Intra-body Communications for Body Area Network Applications // IEEE Transactions on Biomedical Engineering, vol. 60, no. 8, p.2067-2079, 2013.
6. *Kumar S., Nilsen W., Abernethy A., Atienza A., Patrick K., Pavel M., Riley W., Shar A., Spring B., Spruijt-Metz D., Hedeker D., Honavar V., Kravitz R., Lefebvre R., Mohr D., Murphy S., Quinn C., Shusterman V. and Swendem D.* Mobile Health Technology Evaluation: The mHealth Evidence Workshop, American Journal of Preventive Medicine, vol. 45, no. 2, p.228-236, 2013.
7. *Yang K., Chopra N., Upton J., Hao Y., Philpott M., and Alomainy A., Abbasi Q.H and Qaraqe K.* Characterising Skin-Based Nano-Networks for Healthcare Monitoring Applications at THz, 2015 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting p. 199 - 200
8. *Gopinath S., Tang T., Chen Y., Citartan M. and Lakshmipriya T.* Bacterial detection: From microscope to smartphone // Biosensors and Bioelectronics, vol. 60, p.332-342, 2014.
9. *Brufau J., Puig-Vidal M., Lopez-Sanchez J., Samitie J., Driesen W., Breguet J.M.* MICRON: Small Autonomous Robot for Cell Manipulation Applications, Proceedings of the 2005 IEEE International Conference on Robotics and Automation p.844-849.
10. *Gu Z., Aimetti A., Wang Q., Dang T., Zhang Y., Veisoh O., Cheng H., Langer R. and Anderson D.* Biomedical Robotics and Biomechanics Injectable Nano-Network for Glucose- Mediated Insulin Delivery, ACS Nano, 2013, 7 (5), p.4194-4201.

11. *Cho S., Park S. J., Choi Y. J., Jung H., Zheng S., Ko S. Y., Park J. and Park S.* Biomedical Robotics and Biomechanics / 5th IEEE RAS & EMBS International Conference, p.856-860.
12. *Najah A. A., Mervat A.* Internet of NanoThings Healthcare Applications: Requirements, Opportunities, and Challenges // Wireless and Mobile Computing, Networking and Communications (WiMob) p. 9-14, 2015.
13. *Chornaya D., Paramonov A., Koucheryavy A.* Investigation of Machine-to-Machine Traffic Generated By Mobile Terminals //Special Session on Recent Advances in Broadband Access Networks 2014. p. 210 - 213.
14. *Ruslan Kirichek, Margarita Golubeva, Vyacheslav Kulik , Andrey Koucheryavy* The home network traffic models investigation, 2016 18th International Conference on Advanced Communication Technology (ICACT).
15. *Rupp M., Laner M., Svoboda P.* Detecting M2M traffic in mobile cellular networks, IWSSIP 2014 Proceedings, p.159–162, 2014.
16. *Obada Al-Khatib, Wibowo Hardjawana, and Branka Vucetic* Traffic Modeling for Machine-to-Machine (M2M) Last Mile Wireless Access Networks, Globecom 2014 - Communications QoS, Reliability and Modelling Symposium. p. 1199 - 1204
17. *Paramonov A., Koucheryavy A.* M2M traffic models and flow types in case of mass event detection, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), p. 294-300, 2014.
18. *Parka, K.* On the Effect of Traffic Self-similarity on Network Performance / K. Parka, G.Kimb, and M.Crovellab // In Proceedings of the SPIE International Conference on Performance and Control of Network Systems.-1997, p.296-310.

УДК 517.8

Моделирование пространственных структур статистической зависимости экстремального типа

Е. Ю. Щетинин, В. Мерзляков

*Кафедра прикладной математики,
МГТУ Станкин,
Вадковский пер., д.1, Москва, Россия, 127055*

Аннотация. Парадигма копул является интенсивно развивающейся математической теорией моделирования структур зависимости многомерных случайных величин. Немалое место в ней занимают структуры статистической зависимости экстремального типа. Общий теоретический подход к конструированию моделей экстремального типа зависимости известен и уже достаточно хорошо описан в различных работах, однако такому разделу как разработка моделей копул экстремального типа на наш взгляд уделено недостаточно места в общем потоке публикаций. В работе рассмотрены основные положения теории копул, многомерных структур экстремального типа, а также их взаимосвязи. Далее мы обобщили предложенную копулу на случай пространственных связей переменных и разработали численный алгоритм оценивания ее параметров. Рассмотрены приложения пространственных моделей копул экстремального типа для описания и анализа поведения таких экстремальных событий как лесные пожары в пространственной области.

Ключевые слова: копула, экстремальные значения, пространственные области, пространственные структуры зависимости.

1. Введение

В работе рассмотрены основные положения теории копул, многомерных структур экстремального типа, а также их взаимосвязи. Далее мы обобщили предложенную копулу на случай пространственных связей переменных и разработали численный алгоритм оценивания ее параметров. Рассмотрены приложения пространственных моделей копул экстремального типа для описания и анализа поведения таких экстремальных событий как лесные пожары в пространственной области.

2. Основная часть

2.1. Моделирование структур статистической зависимости с использованием копул

Математическая теория копул является современным интенсивно развивающимся направлением математического моделирования структур статистической зависимости и представляет собой прекрасный инструментарий по разработке различных моделей распределения k -мерного случайного вектора $Z = (Z_1, \dots, Z_k)$ [1, 2]

$$Pr(Z_1 \leq z_1, \dots, Z_k \leq z_k) = C(u_1, \dots, u_k). \quad (1)$$

Функция копулы C определена на единичном кубе $[0, 1]^k$ и обладает частными стандартными равномерными распределениями $u_j = P(Z_j < z_j)$, $j = 1, \dots, k$. Мы сосредоточимся на классе копул, которые характеризуют так называемые структуры зависимости экстремального типа [2]. Функция называется копулой экстремального типа, если существует функция C^* , удовлетворяющая условию

$$C(u_1^{1/n}, \dots, u_k^{1/n}) \rightarrow C^*(u_1, \dots, u_k), n \rightarrow \infty, \quad (2)$$

$(u_1, \dots, u_k) \in [0, 1]^k$. Можно показать, что справедливо следующее равенство [?, 4]

$$C^*(u_1, \dots, u_k)^n = C^*(u_1^n, \dots, u_k^n), n > 0.$$

В работе [7] предложена спектральная характеристика многомерных распределений экстремального типа, которую в терминологии (1)-(2) можно записать в следующем виде

$$C^*(u_1, \dots, u_k) = \exp \left[-V \left(-\frac{1}{\log u_1}, \dots, -\frac{1}{\log u_k} \right) \right], \quad (3)$$

где функция V обладает следующими свойствами $V(nu_1, \dots, nu_k) = n^{-1}V(u_1, \dots, u_k)$ для всех $n > 0$ и называется спектральной экспонентой. В частности, известно представление Пикендса для экстремальных копул

$$C^*(u_1, \dots, u_k) = \exp \left[-A \left(\frac{\log u_1}{\sum_i u_i}, \dots, \frac{\log u_k}{\sum_i u_i} \right) \sum_i u_i \right],$$

где функция A - функция экстремальной зависимости. Весьма распространенным примером является экстремальная копула Демарты-Макнейла [4]. При $k=2$

$$A(w) = wT_{v+1} \left[\frac{\left(\frac{w}{1-w} \right)^{1/v} - \rho}{\left(\frac{1-\rho^2}{1+v} \right)^{1/2}} \right] + (1-w)T_{v+1} \left[\frac{\left(\frac{1-w}{w} \right)^{1/v} - \rho}{\left(\frac{1-\rho^2}{1+v} \right)^{1/2}} \right], \quad (4)$$

$0 < w < 1$, $-1 < \rho < 1$, T_v - функция распределения Стьюдента с v степенями свободы. В пределе при $v \rightarrow \infty$ коэффициент корреляции $\rho = \exp\left(-\frac{a^2}{2v}\right) \sim 1 - \frac{a^2}{2v}$ для некоторого $a \in [0, \infty]$, выражение (3) переходит в копулу Хеслера-Рейсса

$$A(w) = (1-w)\Phi\left[\frac{a}{2} + a^{-1}\log\left(\frac{1-w}{w}\right)\right] + w\Phi\left[\frac{a}{2} + a^{-1}\log\left(\frac{w}{1-w}\right)\right]. \quad (5)$$

Здесь Φ - стандартная нормальная функция распределения. С другой стороны, предел выражения (4) при $v \rightarrow \infty$ получим копулу Маршалла-Олкина

$$C(u_1, u_2) = \exp(\alpha \log(u_1 u_2) + (1 - \alpha) \log(\min(u_1, u_2))), 0 \leq \alpha \leq 1.$$

Единственной архимедовой копулой среди копул экстремального типа является логистическая копула Гумбеля-Хогарда

$$C^*(u_1, \dots, u_k) = \exp\left[-\left(\sum_{j=1}^k (-\log u_j)^{1/\alpha}\right)^\alpha\right], 0 < \alpha \leq 1.$$

В качестве важного примера приведем также копулу Галамбоша, известную как отрицательно логистическая копула

$$C^*(u_1, \dots, u_k) = \exp\left[-\sum_{\substack{J \subset \{1, \dots, k\} \\ |J| \geq 2}} (-1)^{|J|} \left(\sum_{j \in J} (-\log u_j)^{-\alpha}\right)^{-1/\alpha}\right] \prod_{j=1}^k u_j,$$

где $\alpha > 0$. Здесь внешнее суммирование производится по всем подмножествам J порядка $\{1, \dots, k\}$, мощность которых $|J|$ больше чем 2.

2.2. Моделирование пространственных структур статистических зависимостей

В этом разделе мы попытались расширить представление (3) на пространственный случай, когда переменные становятся случайными процессами, наблюдаемыми на некоторой области X в различных локациях $x_1, \dots, x_d \in X, X \subset \mathbb{R}^d$. Основной проблемой становится невозможность расширения некоторых классов копул на пространственный случай, а также сложности, возникающие при оценивании параметров моделей копул, как это показано в работе [6]. Несомненно, привлекательными в этой ситуации становятся копулы (4)-(5), допускающие обобщение на бесконечномерный случай и характеризующиеся известными корреляционными функциями. Например, соответствующее

расширение для гауссовского процесса может выглядеть следующим образом

$$Z(x) = F_x^{-1}[\Phi(\varepsilon(x))], x \in X,$$

где $\varepsilon(x)$ — белый шум, F_x^{-1} — обобщённая обратная функция частного распределения случайной величины $Z(x)$ в локации x . Наибольший интерес для нас представляет пространственное расширение обобщенного распределения экстремальных величин

$$F_x(z) = \begin{cases} \exp \left[- \left(1 + \xi(x) \frac{z - \mu(x)}{\sigma(x)} \right)_+^{-1/\xi(x)} \right], \xi(x) \neq 0, \\ \exp \left[- \exp \left(- \frac{z - \mu(x)}{\sigma(x)} \right) \right], \xi(x) = 0, \end{cases}$$

где $x \in X$, u_+ обозначает $\max(u, 0)$, $\mu(x)$, $\sigma(x)$, $\xi(x)$ являются примерами распределения (7).

Фундаментальную роль в описании процесса $Z(x)$ играют максимум-устойчивые процессы, введенные в работе [4]. Пусть $\{S_j^{-1}\}$, $j = 1, \dots, \infty$ являются значениями пуассоновского процесса на \mathbb{R}^+ с интенсивностью $d\Lambda(s) = s^{-2}ds$ и пусть $(W_j(x))_{j=1}^\infty$ — независимые копии стационарного процесса $W(x)$ на \mathbb{R}^d , удовлетворяющего условию

$$\mathbb{E}[\max(0, W_j(0))] = 1.$$

Тогда

$$Z(x) = \max_j S_j \max\{0, W_j(x)\} \quad (6)$$

является максимум-устойчивым стационарным процессом на \mathbb{R} со стандартными частными распределениями Фреше $P(Z(x) \leq z) = \exp(-1/z)$, $Z > 0$. в локациях $x \in X$. Выбор модели процесса $W(x)$ определяет соответствующие модели совместного конечномерного распределения значений $\{Z(x)\}$ в локациях $x_1, \dots, x_d \in X$. Его общее выражение получено в работе [6]

$$P(Z(x_1) \leq z_1, \dots, Z(x_d) \leq z_d) = \exp[-V(z_1, \dots, z_d)] \quad (7)$$

для всех $z_1, \dots, z_d > 0$, с экспоненциальной функцией

$$V(z_1, \dots, z_d) = \mathbb{E} \left\{ \max_{j=1, \dots, d} \frac{W(x_j)}{z_j} \right\}. \quad (8)$$

Тогда, соответствующая функция экстремальной копулы принимает вид

$$C^*(u_1^n, \dots, u_d^n) = \exp \left[\mathbb{E} \left(\max_{j=1, \dots, k} W(x_j) \log u_j \right) \right]$$

$u_1, \dots, u_d > 0, u_j = \exp(-1/z_j), j = 1, \dots, d$. Нами доказано, что копула (8) удовлетворяет условию экстремальности

$$C^*(u_1^n, \dots, u_d^n) = \exp \left[n \mathbb{E} \left(\max_{j=1, \dots, k} Y(x_j) \log(u_j) \right) \right] = C^*(u_1, \dots, u_d)^n$$

В случае, когда $W(x)$ является стационарным гауссовским процессом с корреляционной функцией $\rho(h)$, М.Шлатер показал [5], что функция (8) в двухмерном случае имеет вид

$$V(z_1, z_2) = \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - \frac{2(1 + \rho(h))z_1z_2}{(z_1 + z_2)^2}} \right)^{1/2} \quad (9)$$

Замечательным свойством модели (9) является возможность использования разнообразных корреляционных функций $\rho(h)$. Для частного случая модели (6) $W(x) = \sqrt{2\pi} \max\{0, \varepsilon(x)\}$, где $\varepsilon(x)$ — стандартный гауссовский процесс с семивариограммой $\gamma(h)$ и $\varepsilon(0) = 0$, в работе получено выражение функции копулы для двух локаций $x_1, x_2 \in X$, где $h \in R^+$ — евклидово расстояние $\|x_1 - x_2\|$

$$C^*(u_1, u_2) = \exp \left(\frac{\log(u_1) + \log(u_2)}{2} \right) \left(1 + \sqrt{1 - \frac{2(1 + \rho) \log(u_1) \log(u_2)}{(\log(u_1) + \log(u_2))^2}} \right),$$

$$-1 \leq \rho \leq 1.$$

Полагая $W(x) = \exp(\varepsilon(x) - \gamma(x))$, получим модель Брауна-Резника [9], если принять $\gamma(h) = h^\alpha, 0 < \alpha < 2$. При $\alpha = 2$ получим модель Смита [8]

$$V(z_1, z_2) = \frac{1}{z_1} \Phi \left(\frac{\alpha(h)}{2} + \alpha^{-1}(h) \log \left(\frac{z_2}{z_1} \right) \right) + \Phi \left(\frac{\alpha(h)}{2} + \alpha^{-1}(h) \log \left(\frac{z_2}{z_2} \right) \right) \quad (10)$$

где $\alpha^2(h) = h^T \Omega^{-1} h$ — расстояние Махаланобиса.

В качестве меры статистической зависимости двух пространственных максимум-устойчивых процессов в двух локациях $x_1, x_2 \in X$ часто используют коэффициент экстремальной зависимости $\theta(h) = V(1, 1) \in [1, 2]$. При этом значение $\theta(h) = 1$ соответствует полной зависимости, $\theta(h) = 2$ соответствует независимости. Коэффициент $\theta(h)$ представляет собой весьма полезный инструментарий для анализа статистических связей пространственных случайных процессов. Так, известно, что корреляционная функция гауссовского процесса $\rho(h) \rightarrow 0$ при $h \rightarrow \partial$. Тогда легко показать, что $\theta(h) = 1 + 2^{-1/2} \approx 1.707$, т.е. в рамках модели (10) независимые экстремумы на больших расстояниях практически не возникают.

3. Заключение

Таким образом, в настоящей работе мы кратко рассмотрели основные положения теории экстремальных величин и ее расширение на пространственные случайные процессы. В качестве математической модели пространственных процессов было предложено использовать процессы устойчивых максимумов, краткое описание которых также приведено в работе. Были представлены некоторые известные модели пространственных максимум-устойчивых процессов, для них были получены соответствующие выражения совместных распределений для двух различных локаций пространственной области. Решение актуальных задач геостатистики сделало востребованными модели копул экстремального типа. В работе построены модели пространственных структур статистической зависимости экстремального типа на основе расширения копул на пространственную область и использования моделей Шлатера и Резника-Брауна.

Литература

1. В.А. Акимов, Быков В.А., Щетинин Е.Ю. Введение в статистику экстремальных значений и ее приложения, МЧС России. — М.:ФГУ ВНИИ ГОЧС(ФЦ).2009.-524с.
2. H. Joe. Multivariate models and dependence concepts. Chapman & Hall, London, 1997.
3. J.A. Tawn. Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3):397–415, 1988.
4. L. de Haan. A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204, 1984.
5. M. Sclater, J. A. Tawn. A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika* 90, 139–156, 2003.
6. A. C. Davison, M. M. Gholamrezaee. Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 10, 2011.
7. L. de Haan, S. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40:317–337, 1977.
8. B. M. Brown, S. I. Resnick. Extreme values of independent stochastic processes. *J. Appl. Prob.*, 14:732–739, 1977.
9. Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Prob.*, 37(5):2042–2065, 2009.

UDC 517.8

Statistical extreme type dependence structures modeling in spatial domains

E. U. Schetinin, V. Merzlyakov

*Department of Applied Mathematics
Moscow State Technological University "Stankin"
Vadkovskiy sidestr. 1, Moscow, Russia, 127055*

The paradigm of the copula is an intensively developing theory of mathematical modeling of structures of dependence of multidimensional random variables. The considerable part are structures of statistical dependence of the extreme type. A General theoretical approach to the construction of models of extreme type of dependence is known and sufficiently well described. However, such section as modeling copulas extreme type in our view paid insufficient space in the general flow of publications.

Keywords: structures dependence, copula, extreme type, spatial domain.

УДК 004.71

Влияние трафика Интернета Вещей на работу сетевого оборудования

А. А. Серебрякова, В. А. Кулик, В. Д. Фам, Р. В. Киричек

** Кафедра сетей связи и передачи данных,
Санкт-Петербургский университет телекоммуникаций им. проф. М.А.
Бонч-Бруевича,
пр. Большевиков, д.22, Санкт-Петербург, Россия, 193232*

Аннотация. Концепции, связанные с "Будущим Интернета" исследуются, разрабатываются и постоянно адаптируются к повседневной жизни. Одна новая концепция, связанная с "Будущим Интернета" называется "Интернет Вещей" (Internet of Things, IoT). Эта концепция предполагает, что реальные объекты являются частью Интернета: каждый объект уникально идентифицируется, и доступен в сети, его положение и статус известны, многочисленные услуги и новостная информация приумножаются, чтобы расширить Интернет возможности, легко комбинируя между цифровым и физическим миром, в конечном счете Интернет Вещей влияет на личную и социальную жизнь. На протяжении нескольких лет Интернет Вещей стремительно разрастался и темпы роста только увеличиваются, и как следствие, наблюдается увеличение объемов трафика Интернета Вещей. Статья посвящена исследованию влияния трафика Интернета Вещей на работу сетевого оборудования, эта задача представляет собой одну из приоритетных в области телекоммуникаций, так как этот вид трафика на сегодняшний день повсеместно используется.

Ключевые слова: Интернет Вещей, процессор, оперативная память, загрузка.

1. Введение

При бурном технологическом развитии, приводящем к тому, что растет быстродействие компьютерных систем и коммуникационных ресурсов и систем в целом, количество пользователей стремительно растёт. Исследования, проведенные в Китае в период с 2001 по 2006 год, показали, что количество объема трафика растет подобно закону Мура, т. е. удваивается каждые 5,32 года [1]. Можно предположить, что примерно через 5 лет нас может затронуть проблема с использованием сетевого оборудования. Возможно, что, то сетевое оборудование, которое используется сейчас, через небольшой промежуток времени не сможет стабильно функционировать на сети. Все это можно наблюдать по причине того, что люди помимо того, что применяют сети связи в своей профессиональной деятельности (ip-телефония, удаленная работа, дистанционное обучение и т. д.) и просто для досуга (чаты, web, игры, музыка, видео и т. д.). Они хотят контролировать все, что их окружает. Хотят знать, что у них творится дома, что будет с погодой через 5 минут, нет ли пробки по пути домой, также люди

хотят контролировать состояние их организма и многое другое, и в этом им помогает Интернет Вещей. Что немаловажно, Интернет Вещей позволяет предупредить происшествия, например, пожар, ДТП, пробку [2–4, 6]. И как следствие, список востребованных услуг с привлечением Интернета Вещей и загруженность сетевого оборудования растут стремительными темпами. Технологическое развитие не успевает за пользовательскими потребностями, и все чаще случаются ситуации перегруженности сетевого оборудования, и это приводит к задержкам при передаче информации, а иногда к её потерям.

Чтобы найти компромисс между растущими потребностями человечества в ресурсах сетей связи и их предельными возможностями, необходимо использовать специальные алгоритмы управления трафиком и регулирование информационных потоков, а также, что очень важно, определить, какое сетевое оборудование справится с конкретной нагрузкой, а какое нет. Ввиду этого сегодня уже иначе определяются задачи использования телекоммуникационного оборудования, возникает проблема перегрузки сетевого оборудования, и как следствие, необходимость в определении возможностей оборудования. Резко возрастает объем сообщений между устройствами (Machine-to-Machine) [5], уплотняется и усложняется динамика трафика Интернета Вещей. Поэтому актуальными становятся задачи исследования загрузки оборудования, вызванных колоссальным объемом устройств и сервисов в сетях, изучения особенностей трафика Интернета Вещей, а также поиска решения проблемы [7].

2. Основная часть

В настоящее время отсутствуют методики для тестирования сетевого оборудования на устойчивость к трафику Интернет Вещей [8]. Поэтому в этой статье необходимо провести исследование для изучения вопроса о функционировании сетевого оборудования в условиях перегрузок, т.е. провести нагрузочное тестирование. С помощью такого тестирования можно оценить качество работы оборудования и определить, насколько оно устойчиво к конкретным нагрузкам. Сначала проводится тестирование в отсутствии загрузки канала на участке сети, после тестирование будет проводиться при загруженном канале. При этом количество Интернет Вещей, нагружаемых оборудование, будут увеличиваться и при этом отслеживаются следующие параметры: загрузка процессора и оперативной памяти.

Исследуются беспроводные маршрутизаторы марки ASUS моделей RT-N11P и WL-500gP V2, также марки MikroTik модели hAP lite (RB941-2nD-TC). Так как исследование посвящено тестированию беспроводного маршрутизатора в условиях перегрузок, нужно загрузить трафиком Интернета Вещей, в нашем случае это трафик протокола MQTT, центральный процессор и оперативную память, и отслеживать эти параметры. Тестирование беспроводной точки доступа осуществляется с помощью управляющего компьютера, на котором запускается

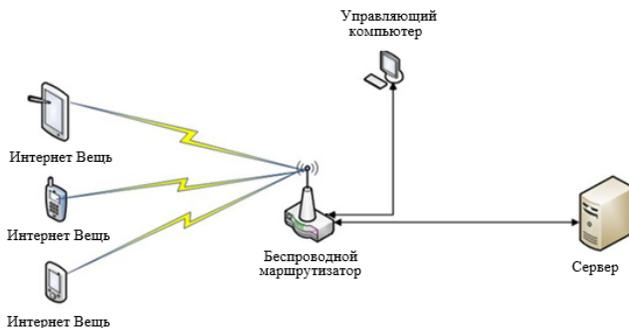


Рис. 1. Измерение загрузки процессора и оперативной памяти маршрутизатора

процесс передачи пакетов, таким образом широковещательно нагружая маршрутизатор трафиком от Интернет Вещей, с помощью него можно изменять трафик, вносить в него задержку, выбирать протоколы, которые необходимы в исследовании, так же на нем можно включить генератор трафика. Благодаря веб интерфейсу маршрутизатора, можно отслеживать измерения значений основных характеристик.

В качестве Интернет Вещей использовались: - Микрокомпьютеры (Intel Edison, Intel Galileo); - Микроконтроллеры с ограниченной вычислительной мощностью (ESP8266).

Для проведения эксперимента была собрана схема на базе модельной сети лаборатории Интернет Вещей Санкт-Петербургского университета телекоммуникаций им.М.А.Бонч-Бруевича [9], согласно рис. ?? и использовано оборудование, о котором сказано выше. Был проведён эксперимент, результаты которого отображены на графиках (рис.2 и рис.3).

На графиках на рис. ?? проиллюстрированы 4 теста, проведенные с тремя маршрутизаторами. Первый тест – нижний график был проведен без нагрузки. Второй – график выше – маршрутизатор нагружается 12-ью Интернет Вещами. Третий – в пределах от 80% до 100% - подключено 24 Интернет Вещи. Четвертый – колеблется в районе 100% - используется генератор трафика, причем такая загрузка достигается при отсутствии Интернета Вещей. Загрузка оперативной памяти при этом составила в среднем 24,34 МБ из 32 МБ, 12,22 МБ из 16 МБ и 24,12 МБ из 32 МБ для ASUS RT-N11P, ASUS WL-500gP V2 и MikroTik hAP lite (RB941-2nD-TC), соответственно. Так же в эксперименте был момент, когда Интернет Вещи не могли подключиться к серверу и слали только служебные сообщения, тогда можно было наблюдать загрузку процессора до 20 - 30%.

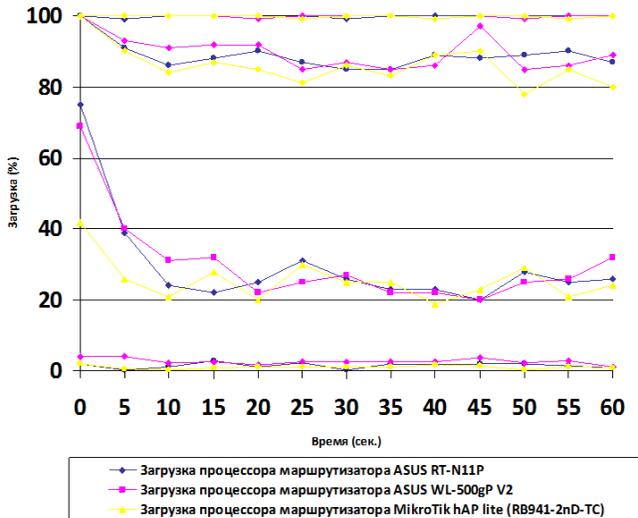


Рис. 2. Измерение параметров с применением протокола MQTT

Так же можно отметить, что трафик MQTT схож с трафиком UDP и отличается от TCP трафика. Это видно на графиках, изображенных на рис. ??

3. Заключение

Эксперимент показал, что трафик Интернета Вещей значительно влияет на загрузку процессора и оперативной памяти. Загрузка процессора напрямую зависит от количества, подключенных Интернет Вещей, а также от индивидуальных алгоритмов обработки потоков данных и мощности маршрутизаторов.

При таких прогнозах актуально говорить о переходе на программно-определяемые сети, т. к. с их применением возможно управление трафиком, распределение политик безопасности в масштабе всей сети. SDN может помочь отделить или блокировать сетевые маршруты, где обнаружена атака или нарушение безопасности, эти сети позволяют значительно автоматизировать и облегчить управление Интернет Вещами за счет возможности их гибкого конфигурирования.

Полученные результаты планируется использовать для выбора шлюза для медицинских сетей.

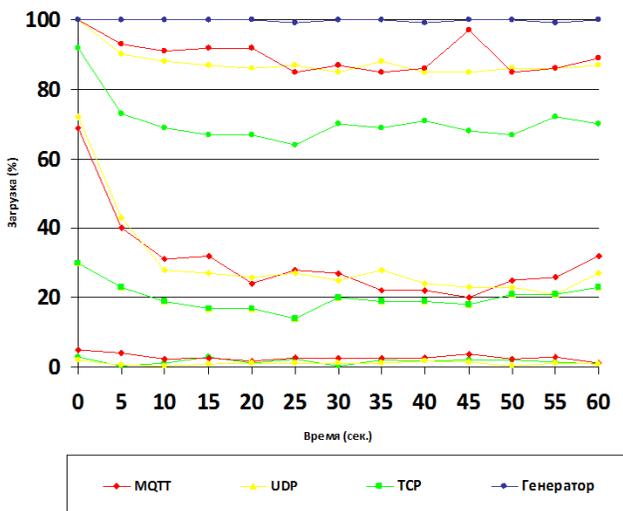


Рис. 3. Измерение параметров с использованием протоколов MQTT, UDP и TCP

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта N16-37-00215 «Биодрайвер».

Литература

1. А. В. Росляков, С. В. Ваняшин, А.Ю. Гребечников, М. Ю. Самсонов Интернет Вещей - Самара, 2014.
2. А.Е.Кучерявый Интернет Вещей // Электросвязь. - 2013. - № 1. с.21-24.
3. А.Е.Кучерявый, А.Е.Прокопьев, Е.А.Кучерявый Самоорганизующиеся сети. - СПб.: Любавич, 2011. - 312 с.
4. Курчечек Р. В., Парамонов А. И., Прокопьев А. В., Кучерявый А. Е. Эволюция исследований в области беспроводных сенсорных сетей // Информационные технологии и телекоммуникации. — 2014. — № 4 (8). — С. 29-41.
5. Y.2069. Terms and definitions for the Internet of things // Recommendation ITU-T // Jul 2012.
6. А. Е. Кучерявый, П. Н. Боронин Интернет Вещей как новая концепция развития сетей связи // Информационные технологии и телекоммуникации. -2014. - №3 (7). - С. 7-30.

7. *Парамонов А. И.* Разработка и исследование комплекса моделей трафика для сетей связи общего пользования // Диссертационная работа
8. *Васильев А.Б., Тарасов Д.В., Андреев Д.В., Кучерявый А.Е.* Тестирование сетей связи следующего поколения. М.:ФГУП ЦНИИС, 2008 г. – 144 с.
9. *Kirichek R., Koucheryavy A.* Internet of Things Laboratory Test Bed // Lecture Notes in Electrical Engineering. 2016. Vol. 348. PP. 485-494.

UDC 004.71

Effect of Traffic IoT on network equipment

A. A. Serebryakova, V. A. Kulik, V. D. Pham, R. V. Kirichek

** Department of Communications Networks and Data Transmission
The Bonch-Bruевич Saint Petersburg State University of Telecommunications
Bolshevikov Ave., h.22, St. Petersburg, Russia, 193232*

Concepts associated with the “Future Internet” are being researched, developed and continuously adapted in daily life. One new concept associated with the “Future Internet” is called “Internet of Things” (IoT). The IoT become a vision where real-world objects are part of the internet: every object is uniquely identified, and accessible to the network, its position and status known, where numerous services and intelligence are added to effectively expand an Internet, seamlessly combining between the digital and physical world, eventually affecting on personal and social environment. For several years, the Internet of Things grew rapidly and the growth rate is only increasing, and as a consequence, one can see an increase in volume of Internet of Things traffic. The article investigates the impact of the Internet of Things traffic on network equipment, this task is one of the priorities in the field of telecommunications, since this type of traffic to date commonly used.

Keywords: Internet of Things, processor, random access memory, utilization.

УДК 004.725.4:355.469.34

Исследование возможности применения роя БПЛА для организации инфраструктуры сети VANET

П. А. Шилин, Р. В. Киричек

*Кафедра сетей связи и передачи данных,
Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М. А. Бонч-Бруевича,
пр.Большевиков д.22, корп.1, Россия, Санкт-Петербург, 193232*

Аннотация. В данной работе предлагается рассмотреть возможность использования группы беспилотных летательных аппаратов для беспроводных мобильных самоорганизующихся сетей VANET. Сами сети VANET, используемые для построения интеллектуальных транспортных сетей и систем, в основном предназначенных для обеспечения безопасного движения автомобилей. Для таких сетей характерна быстро меняющаяся топология, быстро движущиеся относительно друг друга узлы, а как следствие невозможность создать надежную для передачи данных сеть на длительное время. Применение БПЛА для данной сети позволяет создать новый вид инфраструктурных узлов, упростить топологию и повысить качество обслуживания клиентов.

Ключевые слова: VANET, БПЛА, IEEE 802.11p, ИТС.

1. Введение

Последние несколько лет, значительно выросло количество работ в области применения беспилотных летательных аппаратов (БПЛА) как временных узлов сети, ретрансляторов, сборщиков данных с последующей доставкой как для военных, так и гражданских применений [1]. У БПЛА, как у любого устройства, есть как преимущества, так и недостатки. Основным преимуществом считается возможность быстрого развёртывания инфраструктуры сети. Из недостатков отмечается время автономной работы, которое увеличивается с появлением новых технологий энергоснабжения БПЛА. Поэтому в научных работах исследователи стали предлагать применять БПЛА для различных сетей связи [2], в том числе, для сетей автотранспорта VANET [3]. В свою очередь, в рамках применения VANET [4], имеется большое количество научных работ, предлагающих различные приложения взаимодействия автомобилей. Появление новых и развитие уже принятых стандартов автомобильного беспроводного взаимодействия, таких как DSRC (Dedicated Short-Range Communications) [5], IEEE 802.11p (Wireless Access in Vehicular Environment, WAVE [6, 7]), C2C [8] усилило эффективность и целесообразность автомобильных коммуникаций.

Таблица 1

Сравнение преимуществ систем с одним БПЛА и множеством БПЛА

Критерий сравнения	Один БПЛА	Множество БПЛА
Вероятность сбоя	Высокая	Низкая
Масштабируемость	Ограниченная	Высокая
Жизнеспособность системы	Низкая	Высокая
Вероятность потери упр.	Высокая	Низкая
Сложность управления	Низкая	Высокая
Стоимость решения	Средняя	Высокая

До недавнего времени, при создании летающих сетей связи использовался один БПЛА. В таких сетях как правило этот БПЛА обслуживал один и несколько наземных узлов. Основными проблемами одиночного летающего узла является: ограниченное радиуса взаимодействия с наземными узлами, предел числа узлов сети, возможного для одновременного слушания. Хорошим вектором развития является использование БПЛА в виде роевых структур или нескольких БПЛА систем, которые объединяясь в единую сеть выполняют задачи массового обслуживания наземных узлов. Работая скоординировано, они могут осуществлять расширение зоны покрытия, организовать систему ретрансляции и др. В зависимости от приложения или услуги меняется степень мобильности БПЛА [9]. Недостатком такой системы является то, что БПЛА иногда могут выходить из строя из-за разряда батареи или неисправности. В таблице 1 мы привели краткий обзор характеристик использования моно и мульти-беспилотных систем. По сравнению с системой Все эти недостатки вероятнее всего уже можно компенсировать, используя протоколы для сетей VANET, так как, в VANET уже реализовано множество подходов, протоколов, которые функционируют достаточно эффективно в условиях постоянных разрывов, быстрых изменений топологии, высокой скорости перемещения узлов.

Поэтому в данной работе мы хотим провести анализ возможности использования роевых структур БПЛА для сетей VANET.

2. Основная часть

Рассмотрим задачу организации сети связи VANET с использованием мульти-беспилотного подхода где каждый БПЛА представляет собой временный узел связи. За основу мы возьмём сеть стандарта IEEE 802.11p, так как именно этот стандарт специально предназначен для построения сетей автотранспорта. Участниками сети могут являться как автомобили, так и БПЛА, способные обмениваться данными

между собой. Находясь в зоне радиодоступности друг друга, автомобили и БПЛА самостоятельно организуют канал связи и используют его для передачи данных. Чтобы возможность применения такой сети связи, и оценить ее эффективность, мы должны сравнить между собой мультикоптерную и монокоптерную системы, представленные на рисунке 1 и 2.

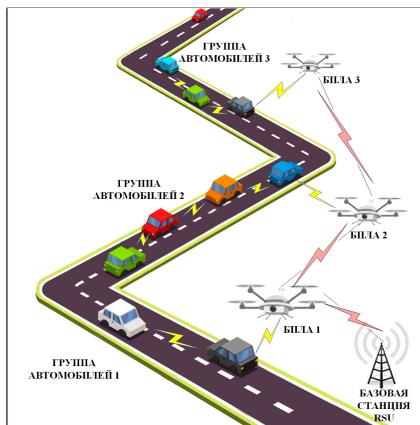


Рис. 1. Мультикоптерная система

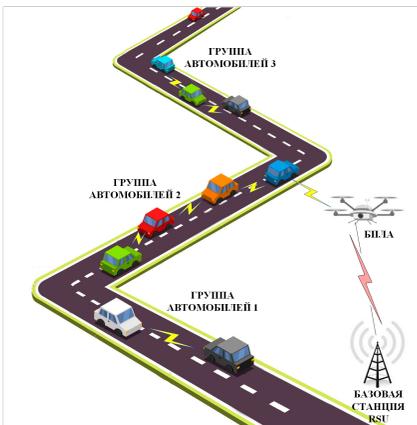


Рис. 2. Монокоптерная система

В качестве узла назначения данных, куда мы должны поступать все данные от различных узлов, необходимо выбрать RSU (Road State Unit), базовую станцию, предоставляющую некоторый сервис для автомобилей, как это рекомендуется при построении архитектуры V2I. В качестве критериев оценки мы будем использовать распределение процента доставленных пакетов для каждого транспортного средства, процент времени нахождения, быстрота времени доставки данных.

Немаловажным элементом для получения максимально точных данных для оценки эффективности системы является модель перемещения узлов. Модель перемещения — это набор функций, описывающих изменение местоположения, скорости, ускорения с течением времени для каждого узла. Модели перемещений используются для создания реалистичного моделирования в окружающей среде. Основными видами моделей являются: модель случайных точек, модель определенных точек, Гаус-Марковская модель и другие, предлагаемые научным сообществом. В рамках данной работы необходимо применить две модели перемещения, как для автомобилей, так и для БПЛА. Допустим, что транспортные средства двигаются по дороге из точки A в точку B или наоборот, тогда модель перемещения для автомобилей будет описывать как множество точек на отрезке от A до B , поэтому

Таблица 2

Параметры симуляции

Стандарт	IEEE 802.11p
Частота канала службы	5870 МГц
Мощность передатчика	200 мВт/23.0 дБ
Чувствительность приемника	-95 дБм
Скорость соединения	6 Мбит/с
Минимальный интервал генерации данных	0.2 сек
Количество автомобилей в симуляции	30 шт
Количество БПЛА в симуляции	1-3 шт

при моделировании движения транспортных средств достаточно выбрать N точек на отрезке AB и в этих точках осуществить сеанс связи и доступными узлами. Для БПЛА существует также большой набор возможных моделей перемещения, но для решения нашей задачи мы допустим, что БПЛА находятся в заранее определенных точках и единой высоте. Точки разнесены относительно друг друга на максимальном расстоянии, при котором возможна организация стабильной связи между ними.

3. Моделирование

Для исследования возможности использования мультибеспилотной системы было проведено имитационное моделирование. Для организации процесса моделирования использовался набор программного обеспечения организующего дискретно-событийную среду. Основой системы моделирования является ПО OMNeT++ [10]. Особый компонент Mixim [11], работающий в связке с OMNeT++, позволяет промоделировать беспроводную передачу для автомобильных сетей. Третий компонент системы моделирования - Veins [12], представляет собой полный набор инструментов, позволяющий создать реалистичные автомобильные маршруты на основе топографических карт. Для симуляции мы настроили среду моделирования согласно параметрам, приведенным в таблице 2.

В результате моделирования мы получили данные распределения процента доставленных пакетов для каждого автомобиля, представленного на рис. 3 и 4, и процент времени нахождения автомобиля в связи с узлом БПЛА, для систем с одиночным и множеством БПЛА, рис. 5 и 6.

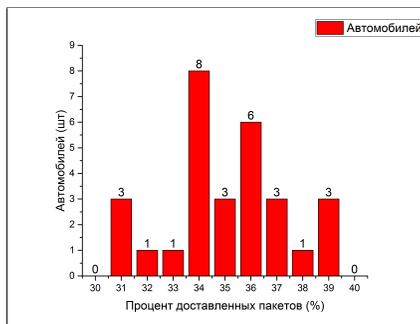


Рис. 3. Распределение доставленных пакетов при одном БПЛА

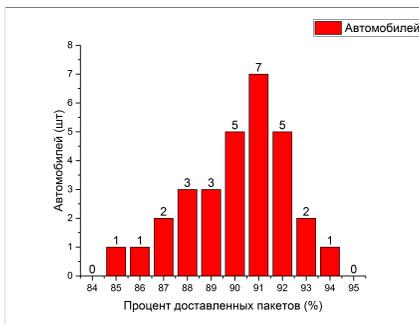


Рис. 4. Распределение доставленных пакетов при роле БПЛА



Рис. 5. Распределение времени радиодоступности наземных узлов при одном БПЛА

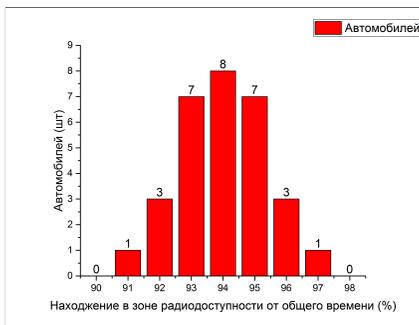


Рис. 6. Распределение времени радиодоступности наземных узлов при роле БПЛА

4. Заключение

Входе проведенного имитационного моделирования и последующего анализа графиков, однозначно можно сказать что система из множества БПЛА работает более эффективнее чем система из одного БПЛА, обеспечивая более высокую вероятность доставки данных и значительно расширяя радиус действия сети, а как следствие время возможного обслуживания узла. Стандарт IEEE 802.11р способен также функционировать на стыке автомобиль-БПЛА. Применение мульти-беспилотных систем для сетей VANET даст значительный прирост к качеству обслуживания каждого узла сети.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-07-09431а «Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей».

Литература

1. *Koucheryavy A., Vladyko A., Kirichek R.* State of the Art and Research Challenges for Public Flying Ubiquitous Sensor Networks Internet of Things, Smart Spaces, and Next Generation Networks and Systems Ed. by S. Balandin, S. Andreev, Y. Koucheryavy. — Springer International Publishing Switzerland. — LNCS. — Vol. 9247. — P. 299–308., 2015.
2. *А.Е.Кучерявский, А.Е.Проконьев, Е.А.Кучерявский* Самоорганизующиеся сети. — СПб.: Любавич, 2011. — 312 с.
3. *Hartenstein H., Laberteaux K.* VANET Vehicular Applications and Inter-Networking Technologies. — WILEY, 2009. — ISBN: 9780470740569.
4. *Kenichi M.* Wide-Area Disaster Surveillance Using Electric Vehicles and Helicopters — 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), — P. 3466–3471, 2013
5. Vehicle Safety Communications Project Final Report U. S. Dept. Trans., Nat. Highway Traffic Safety Admin., Rep. DOT HS 810 591, 2006.
6. IEEE P1609.4-2010 — IEEE Standard for Wireless Access in Vehicular Environments (WAVE), 2009.
7. IEEE 802.11p-2012 — IEEE Standard for Information technology — Local and metropolitan area networks — Specific requirements — Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments, 2012.
8. ETSI TS 102 636–3 V1.1.1 (2010–03): Intelligent Transport Systems (ITS) — Vehicular Communications — GeoNetworking — Part 3: Network architecture. — European Telecommunications Standards Institute, 2010.
9. *Kirichek R., Paramonov A., Koucheryavy A.* Swarm of Public Unmanned Aerial Vehicles as a Queuing Network Communications in Computer and Information Science — Vol. 601. — P. 111-120., 2016.
10. OMNeT++ — Network Simulation Framework. <http://www.omnetpp.org/>.
11. MiXiM —Mixed simulator. <http://mixim.sourceforge.net/>.
12. Veins — open source vehicular network simulation framework. <http://veins.car2x.org/>.

UDC 004.725.4:355.469.34

Research the possibility of using UAVs swarm for organization VANET infrastructure

P. A. Shilin, R. V. Kirichek

*Department communications networks and data transmission,
Saint Petersburg State University of Telecommunications. prof.
Bonch-Bruevich,
pr.Bolshevikov d.22, building 1, Russia, St. Petersburg, 193232*

In this paper we propose to consider the use of unmanned aerial vehicles for mobile wireless self-organizing networks VANET. VANET used to build intelligent transport networks and systems, mostly designed to provide safe movement of vehicles. For such networks is characterized by rapidly changing topology, moving quickly relative to each other nodes, and as a consequence the inability to create a reliable data transmission network. The use of UAVs for this network allows you to create a more reliable nodes, to simplify the topology and improve the quality of customer service.

Keywords: VANET, UAVs, IEEE 802.11p, ITS.

УДК 004.7

Обзор возможных видов и методов тестирования летающих сенсорных сетей

А. В. Шкляева, Р. В. Киричѣк

*Кафедра сетей связи и передачи данных,
Санкт-Петербургский Государственный Университет Телекоммуникаций
им. проф. М. А. Бонч-Бруевича,
пр. Большевикова, д.22, Санкт-Петербург, Россия, 193232*

Аннотация. Летающие сенсорные сети приобретают все большую популярность благодаря использованию БПЛА общего пользования во многих сферах жизнедеятельности. Наиболее важным преимуществом летающих сенсорных сетей является способность обеспечивать связь в реальном времени без необходимости какой-либо инфраструктуры. Применение проводной связи в данных сетях неуместно из-за большой стоимости и массивности монтажных работ, которые не подходят для маленьких и подвижных устройств.

В данной статье будут рассмотрены возможные виды и методы тестирования летающих сенсорных сетей и приведены разработанные тестовые спецификации для исследований.

Ключевые слова: Летающая сенсорная сеть, FUSN, ZigBee, BLE, 6LoWPAN, IoT, БПЛА, квадрокоптер, сенсорный узел, датчик, тестирование.

1. Введение

В современном мире технологии развиваются с невероятной скоростью, и появление таких устройств, как беспилотные летательные аппараты, не обошло сферу инфотелекоммуникаций. На основе БПЛА активно разворачиваются новые сети связи, которые получили название летающих сенсорных сетей [1].

Благодаря БПЛА, оснащённым средствами связи и датчиками, появилась возможность автоматизированного сбора данных с удаленных сенсорных узлов, которые накапливают информацию за время автономной работы. Во многих сферах жизни такие сети просто необходимы, поэтому летающие сенсорные сети уже давно нашли широкое применение в военных целях для наблюдения и размещения объектов, а также активно используются в сельскохозяйственной сфере, комплексах жизнеобеспечения и для организации резервирования каналов связи при чрезвычайных ситуациях [2, 3].

С помощью данных сетей можно заблаговременно предотвратить какие-либо бедствия, постоянно отслеживая и контролируя небезопасные участки посредством квадрокоптеров, которые собирают информацию с установленных датчиков [4] и моментально передают ее в Интернет [5, 6].

Использование БПЛА с высокой мобильностью и ограниченными энергетическими возможностями вносит не только множество возможностей для предоставления новых услуг, но и ряд проблем, одной из которых является разработка необходимых методик тестирования летающих сенсорных сетей.

Особое место в развитии любых технологий занимает проблема разработки новых стандартов для тестирования. Существуют рекомендации Международного союза электросвязи серии Q.3900-Q.4099 [7], которые описывают принципы построения и тестирования модельных сетей. Разработкой методик тестирования для приложений Интернета Вещей занимается 11 Исследовательская Комиссия МСЭ-Т, и для области IoT была разработана рекомендация Q.3950 – «Методика тестирования и архитектура модельной сети для систем и функций идентификации на основе тегов». Она определяет функции и целевые системы, которые необходимо протестировать, и задает модельные сети для тестирования их взаимодействия и соответствия [8].

К сожалению, этого недостаточно, и поэтому существует необходимость в разработке новых методик и стандартов для тестирования нового приложения Интернета Вещей - летающих сенсорных сетей.

2. Виды взаимодействия сегментов летающей сенсорной сети и методы их тестирования

Для того чтобы приступить к разработке методики тестирования, необходимо определить задачи и объекты летающих сенсорных сетей для дальнейших исследований.

Задач по тестированию летающих сенсорных сетей очень много, основными можно считать те, которые составляют жизненный цикл ЛСС.

Он состоит из следующих стадий [9]:

- 1) Инсталляция сенсорных узлов с БПЛА;
- 2) Определение координат сенсорных узлов и построение карты сенсорного поля с помощью БПЛА;
- 3) Сбор данных с сенсорного поля по оптимальным маршрутам;
- 4) Доставка полученных данных в ССОП;
- 5) Подзарядка сенсорных узлов с БПЛА;
- 6) Деинсталляция сенсорной сети.

На каждом из этапов данного цикла предполагается проведение тестирования.

Также необходимо рассмотреть основные возможные сценарии использования БПЛА в летающих сенсорных сетях и виды взаимодействия между сегментами сети, для того чтобы определить возможные виды тестов.

Далее на рисунке 1 приведены возможные сценарии использования БПЛА в беспроводных сетях [10]:

- a) Увеличение зоны покрытия при помощи БПЛА:

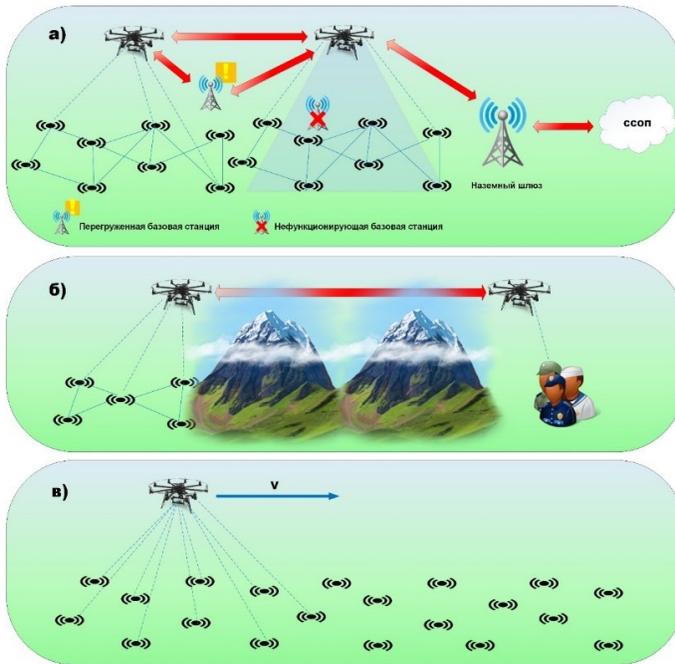


Рис. 1. Сценарии использования БПЛА: а) Увеличение зоны покрытия при помощи БПЛА; б) Ретрансляция информации посредством БПЛА; в) Опрос и сбор данных с сенсорных полей при помощи БПЛА

ЛСС разворачивается для помощи существующей инфраструктуре связи, например, быстрое восстановление связи после частичного или полного повреждения инфраструктуры в результате стихийных бедствий или разгрузка базовой станции в местах с большим скоплением людей (концертные площадки, стадионы и т.д.);

б) Ретрансляция информации посредством БПЛА:

Обеспечение беспроводной связи между двумя и более удаленными пользователями или группами пользователей без надежной сети связи. Это может быть реализовано на пограничных линиях или в центрах экстренного реагирования;

в) Опрос и сбор данных с сенсорных полей при помощи БПЛА:

БПЛА прилетает на сенсорное поле и опрашивает оконечные узлы, получая различные данные (температуру, освещенность, давление, остаточную энергию и т.д.). Примером такого сценария может служить использование ЛСС в сельском хозяйстве.

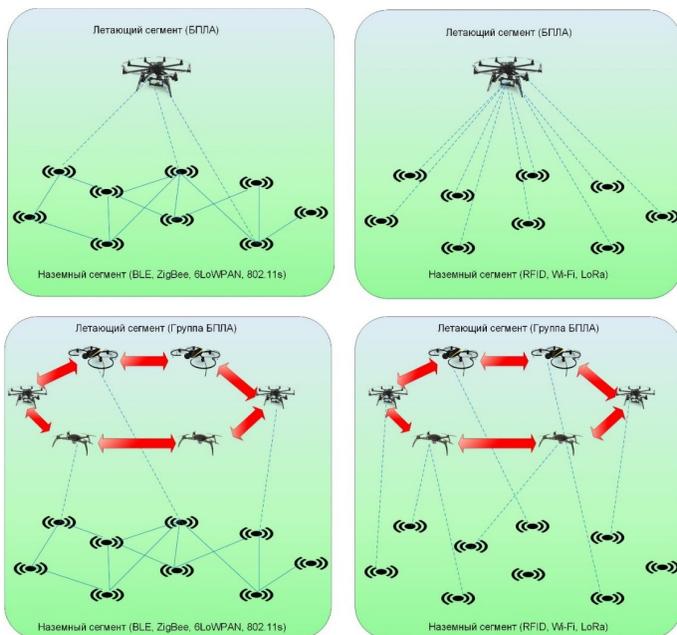


Рис. 2. Виды взаимодействия сегментов ЛСС и возможные технологии построения наземного сегмента

Летающий сегмент сети может быть реализован как на основе одного БПЛА, так и на основе группы БПЛА [11]. Наземный сегмент может быть представлен сетью из самоорганизующихся узлов, поддерживающих технологии (BLE, ZigBee, IEEE 802.11s, 6LoWPAN), а также сетью, которая не поддерживает самоорганизацию узлов (Wi-Fi, RFID, LoRa). Виды взаимодействия сегментов ЛСС показаны на рисунке 2 [12, 13].

Исходя из возможных сценариев использования БПЛА и видов взаимодействия сегментов ЛСС [14], составлена таблица с возможными тестами и методами тестирования относительно используемых сетевых технологий сенсорными узлами [4] наземного сегмента и представлена на рисунке 3.

Далее приведены возможные тестовые спецификации для проведения тестирования сегментов или фрагментов летающей сенсорной сети:

1) Тестирование наземного сегмента ЛСС, построенного на основе протокола ZigBee (определение коэффициента ошибочных пакетов);

	Самоорганизующиеся сети				Несамеоорганизующиеся сети		
	ZigBee	6LoWPAN	802.11s	BLE	Wi-Fi	RFID	LoRa
Однa БПЛА	Физический уровень: 1. Тестирование мощности передатчика; 2. EVM; 3. BER. Канальный уровень: 1. PER; 2. Утилизация канала.	1. Проверка доступности узлов; 2. Проверка совместности; 3. Тестирование производительности.	1. Проверка маршрутов и управления; 2. Проверка целостности сети; 3. Тестирование производительности; 4. Управление ресурсами памяти.	1. Тестирование мощности передатчика; 2. BER; 3. Спектральные измерения с использованием измерительного оборудования (например, Rohde & Schwarz). 4. Измерение параметров модуляции; 5. Измерение частотных характеристик.	Физический уровень: Следует использовать высокоточное измерительное оборудование, позволяющее имитировать ошибки в канале связи (например, Rohde & Schwarz). Канальный уровень: 1. PER; 2. Утилизация канала.	1. Определение чувствительности метки; 2. Определение диаграммы направленности антенны; 3. Проверка совместности устройств между собой.	1. PER; 2. Проверка адаптации скорости передачи данных; 3. Проверка состояния устройства; 4. Проверка возможности регулирования мощности передатчика; 5. Измерение показателей TRP.
	1. Проверка подлинности узлов; 2. Проверка остаточной энергии сенсорных узлов (кроме RFID); 3. Определение периода для опроса и сбора данных с сенсорных узлов (как часто опрашивать?).						
Группа БПЛА	1. Тестирование схожести летающего сегмента при выходе из строя одного или нескольких БПЛА и определение минимального времени для реконфигурации сети: а) если все узлы одинаковы по функциональности; б) если имеется головной узел (ровная структура); 2. Тестирование дальности связи; 3. Проверка подлинности узлов летающего сегмента.						

Рис. 3. Возможные виды тестов относительно используемых сетевых технологий сенсорными узлами наземного сегмента и видов взаимодействия сегментов ЛСС

- 2) Тестирование наземного сегмента ЛСС, построенного по стандарту IEEE 802.11s (проверка целостности сети);
- 3) Тестирование наземного сегмента ЛСС, построенного по технологии LoRa (определение коэффициента ошибочных пакетов);
- 4) Определение периода для опроса и сбора данных с сенсорных узлов;
- 5) Тестирование летающего сегмента ЛСС (тестирование схожести летающего сегмента при выходе из строя одного или нескольких БПЛА и определение минимального времени для реконфигурации сети);
- 6) Тестирование дальности связи.

По предложенным спецификациям предполагается разработка методик тестирования летающих сенсорных сетей.

3. Заключение

В данной статье были рассмотрены основные сценарии использования беспилотных летательных аппаратов в летающих сенсорных сетях, а также виды взаимодействия сегментов ЛСС. Относительно них составлена таблица с возможными видами тестов, на базе которых предложены тестовые спецификации.

В качестве дальнейших исследований предполагается применение разработанных тестовых спецификаций на реальных сетях с использованием производительных модулей и большого числа квадрокоптеров.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №15-07-09431а "Разработка принципов построения и методов самоорганизации для летающих сенсорных сетей".

Литература

1. *Киричек Р. В., Парамонов А. И., Прокопьев А. В., Кучерявый А. Е.* Эволюция исследований в области беспроводных сенсорных сетей // Информационные технологии и телекоммуникации. — 2014. — № 4 (8). — С. 29-41.
2. *Кучерявый А.Е., Владыко А.Г., Киричек Р.В.* Летающие сенсорные сети — новое приложение интернета вещей.// Актуальные проблемы инфотелекоммуникаций в науке и образовании. Международная научно-техническая и научно-методическая конференция: сборник научных статей в 2 томах — СПб. : Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, 2015. — С. 17–22 / РИНЦ.
3. *Кучерявый А. Е., Владыко А. Г., Киричек Р. В., Парамонов А. И., Прокопьев А. В., Богданов И. А., Дорт-Гольц А. А.* Летающие сенсорные сети // Электросвязь. — 2014. — № 9. — С. 2-5.
4. *Y.2069.* Terms and definitions for the Internet of things // Recommendation ITU-T // Jul 2012.
5. *Кучерявый А.Е.* Интернет Вещей // Электросвязь. - 2013. - № 1. с.21-24.
6. *А.Е.Кучерявый, А.Е.Прокопьев, Е.А.Кучерявый* Самоорганизующиеся сети. — СПб.: Любавич, 2011. — 312 с.
7. *Q.3900.* Methods of testing and model network architecture for NGN technical means testing as applied to public telecommunication networks // Recommendation ITU-T // September 2006.
8. *Васильев А.Б., Тарасов Д.В., Андреев Д.В., Кучерявый А.Е.* Тестирование сетей связи следующего поколения. М.:ФГУП ЦНИИС, 2008 г. — 144 с.
9. *Кучерявый А. Е., Владыко А. Г., Киричек Р. В.* Теоретические и практические направления исследований в области летающих сенсорных сетей // Электросвязь. — 2015. — № 7. — С. 9-11.
10. *Zeng Y., Zhang R., Lim T.* Wireless Communications with Unmanned Aerial Vehicles: Opportunities and Challenges. /IEEE Communication Magazin. — 2016 — PP. 36–42.

11. *Koucheryavy A., Vladyko A., Kirichek R.* State of The Art and Research Challenges for Public Flying Ubiquitous Sensor Networks // Lecture Notes in Computer Science. 2015. Vol. 9247. PP. 299-308.
12. *Киричѣк Р. В., Владыко А. Г., Захаров М. В., Кучерявый А. Е.* Модельные сети для Интернета Вещей и программируемых сетей // Информационные технологии и телекоммуникации. — 2015. — № 3 (11). — С. 17-26.
13. *Kirichek R., Kulik V.* Methods of test flying ubiquitous sensor networks // Сборник трудов Международной конференции «Распределенные компьютерные и коммуникационные сети: управление, вычисление, связь 2015». С. 489-499.
14. *Кулик В.А., Киричѣк Р.В., Бондарев А.Н.* Методы исследования беспроводных каналов связи Интернета Вещей в условиях совместной работы // Информационные технологии и телекоммуникации. - 2015. - № 1. - С. 106-114.

UDC 004.7

An Overview of Possible Testing Types and Methods for the Flying Ubiquitous Sensor Networks

A. V. Shklyueva, R. V. Kirichek

*Department of Communication Networks and Data Transmission
The Bonch-Bruевич Saint-Petersburg State University of Telecommunications
pr. Bol'shevikov 22, Saint-Petersburg, 193232, Russia*

Flying sensor networks are becoming increasingly popular due to the use of UAVs in many areas of life. The most important advantage of flying ubiquitous sensor networks is the ability to communicate in real-time without the need of any infrastructure. The use of wire communication in these networks is inappropriate for the high cost and massive installation works, which are not suitable for small and mobile devices.

This article will describe the possible testing types and methods developed for FUSN and also provides the test specifications for research.

Keywords: Flying Ubiquitous Sensor Network, FUSN, ZigBee, BLE, 6LoWPAN, IoT, UAV, quadcopter, sensor node, testing.

УДК 621.39

Об эффективности миграции виртуальных машин в облачных системах

Э.С. Сопин*[†], А. В. Дараселия*, Н.В. Яркина*

* Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198

[†] Институт проблем информатики, ФИЦ ИУ РАН
ул. Вавилова, д.44, Москва, Россия, 119333

Аннотация. Построена модель облачных вычислений в терминах теории массового обслуживания с учетом миграции виртуальных машин и без для сравнения энергопотребления системы.

Ключевые слова: облачные вычисления, энергоэффективность, теория массового обслуживания.

1. Введение

В данной работе рассматривается модель облачной системы Green Cloud, ориентированной на снижение избыточного потребления энергии. Поскольку в среднем простаивающий без нагрузки сервер потребляет от 50% до 70% энергии, затрачиваемой при его полной загрузке, перевод бездействующих серверов в спящий режим позволяет существенно снизить энергопотребление [1]. Потребление энергии сервером в зависимости от его нагрузки может быть описано формулой

$$P(u) = kP_{\max} + u(1 - k)P_{\max},$$

где P_{\max} - максимальная мощность, потребляемая, когда сервер используется полностью, k - доля мощности, потребляемой в режиме ожидания сервера (т.е. 50%), u - доля использования процессора [1]. Для освобождения и выключения серверов с наименьшей нагрузкой применяется механизм миграции виртуальных машин, перераспределяющий их по другим серверам.

2. Анализ эффективности миграции виртуальных машин

Рассмотрим СМО, состоящую из N групп приборов, каждая из которых состоит из V_i приборов. На систему поступает пуассоновский поток заявок с интенсивностью λ , а длительность обслуживания экспоненциальна с интенсивностью μ . Пространство состояний описывается множеством векторов $S = \{ n = (n_1, \dots, n_N) \mid 0 \leq n_i \leq N, 1 \leq i \leq N \}$, где n_i - число занятых приборов в i -ой группе приборов.

При поступлении заявки она идет на обслуживание на наиболее загруженную группу приборов, т.е. $\{n_i < V_i \max(\frac{n_1}{V_1}, \frac{n_2}{V_2}, \dots, \frac{n_N}{V_N}) = \frac{n_i}{V_i}\}$. Если же все приборы заняты, т.е. $n_i = V_i, i = \overline{1, N}$, то заявка теряется.

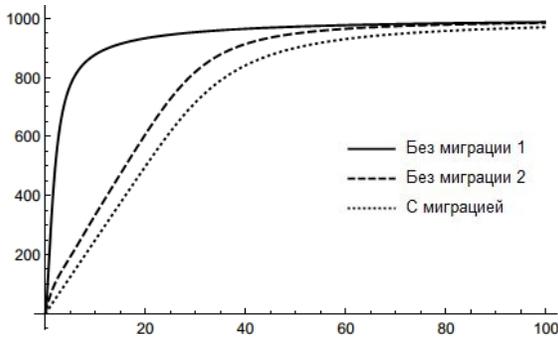


Рис. 1. Графики зависимости мощности P от интенсивности нагрузки λ .

С учетом введенных обозначений СУР имеет вид:

$$p(n_1, \dots, n_N)[(n_1, \dots, n_N)\mu + \lambda] = \sum_{i=1}^N (n_i + 1)\mu p(n_1, \dots, n_i + 1, \dots, n_N) + \lambda \sum_{i=1}^N \varphi_i(n_1, \dots, n_i - 1, \dots, n_N)p(n_1, \dots, n_i - 1, \dots, n_N), \quad n \in S, \text{ где}$$

$$\varphi_i(n_1, \dots, n_N) = \begin{cases} 1, & n_i < V_i; \frac{n_i}{V_i} = \max(\frac{n_k}{V_k}), k = \overline{1, N}; \frac{n_i}{V_i} > \frac{n_j}{V_j}, j < i; \\ 0, & \text{в остальных случаях.} \end{cases}$$

Получив стационарные вероятности системы решая СУР, можно оценить параметры энергопотребления. Средняя суммарная потребляемая мощность вычисляется по формуле [2]:

$$P = \sum_{n \in S} p(n_1, n_2, \dots, n_N) \left[k \sum_{i=1}^N P_i^{\max} \theta(n_i) + (1 - k) \sum_{i=1}^N P_i^{\max} \frac{n_i}{V_i} \right],$$

где P_i^{\max} - максимальная мощность, потребляемая i -м сервером при полной загрузке, $\theta(n_i)$ - функция Хевисайда, т.е.

$$\theta(n_i) = \begin{cases} 0, & n_i = 0, \\ 1, & \text{в остальных случаях.} \end{cases}$$

Проведен численный анализ для сравнения описанной выше модели облачной системы без миграции, модели с равномерным распределением заявок по серверам и модели с мгновенной миграцией. Рассматривается система $N = 4$ серверами, на каждом из которых может быть запущено до 5 виртуальных машин: $V_1 = 5, V_2 = 5, V_3 =$

5, $V_4 = 5$, а интенсивность обслуживания $\mu = 2$. На рисунке 1 представлен график зависимости потребляемой мощности от интенсивности входящего потока при данных исходных параметрах системы.

3. Заключение

Численный анализ показал, что выгода от использования механизмов миграции виртуальных машин на серверах невысока по сравнению со специальными планировщиками, и при некоторых условиях затраты на реализацию этих механизмов становятся существеннее получаемой выгоды.

Благодарности

Исследование выполнено в рамках соглашения № 02.a03.21.0008 от 24.04.2016 г. между Министерством образования и науки Российской Федерации и Российским университетом дружбы народов, а также при частичной финансовой поддержке РФФИ в рамках научных проектов № 14-07-00090, 15-07-03051, 15-07-03608.

Литература

1. *Anton Beloglazov, Jemal Abawajy, Rajkumar Buyya* «Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing» // *Future Generation Computer Systems*, vol.28, 2012. - Pp. 755 - 768.
2. *А.В. Дараселия, Э.С. Сопин* «К вопросу энергоэффективности облачной системы» // *Всероссийская конференция «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем»*, Москва, РУДН, 2016. С. 13–15.

UDC 621.39

On the virtual machines migration effectiveness in cloud systems

E.S. Sopin^{*†}, A. V. Daraseliya^{*}, N. V. Yarkina^{*}

^{*} Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia

[†] Institute of Informatics Problems, FRC CSC RAS,

Vavilova str. 44, 119333 Moscow, Russia

We developed a mathematical model of Green Cloud system with and without live migration of virtual machines in terms of queuing theory for energy efficiency analysis.

Keywords: cloud computing, energy efficiency, queuing theory.

UDC 004.057.4

Per-Packet Load Balancing of TCP Traffic for Goodput Aggregation of Communication Channels with Asymmetric Transmission Delay

V. V. Sosnin*[†]

* *Department of Computer Science,
ITMO University,*

Kronverksky Pr. 49, St. Petersburg, 197101, Russia

[†] *Software Development Department,
Research Institute "Mashtab",*

Kantemirovskaya 5A, St. Petersburg, 194100, Russia

Abstract. Existing methods to aggregate goodput of several channels on a multi-home host require using new protocols such as MPTCP or SCTP. However, in conservative requirements installing or using such protocols may be impossible due to security or budget issues. We explore effectiveness of goodput aggregation using plain old TCP protocol by load-balancing TCP segments on a per-packet weighted round robin basis. Experimental data reveals different network scenarios in which data channels asymmetry affects aggregation efficiency making it possible to determine conditions in which per-packet load-balancing is justifiable.

Keywords: per-packet load-balancing, goodput aggregation, TCP CUBIC, multi-home environment, asymmetric channels.

1. Introduction

Modern computers and even mobile phones can be used as multi-homed hosts nowadays. E.g. a smart-phone may have both 4G and Wi-Fi active Internet connections. By default only one of those connections will be actually used to transmit data, the other channel being reserved for possible fail-over events. But by applying special means, multi-homing may be used to aggregate capacity of several channels and to manage QoS metrics.

By special means we mean specialized protocols such as SCTP or MPTCP ([2]). They both enable effective handling of multi-home environment to aggregate goodput of several channels and fast fail-over recovery. However, those protocols have the following disadvantages:

- The only managed QoS metrics are throughput and loss probability (the last being only partly managed by providing automatic retransmits). Both MPTCP and SCTP are effective for maximizing throughput but not for minimizing RTT and delay variation (jitter in terms of [1]).
- Lack of backward compatibility (both communicating parties have to use new protocols to effectively exploit multi-homing).
- Network programmers have to fix existing programs to get bonuses of multi-homing.

- Some operating systems still have no out-of-the-box support for MPTCP and SCTP, so using them may require additional deployment efforts.

Some authors offer ideas to overcome existing problems. In [5] and [6] new load balancing algorithms are proposed to aggregate goodput of several network channels. Proposed algorithms have the following shortcomings:

- To use the algorithm one needs to modify Linux kernel.
- The algorithm efficiency is not analyzed for different TCP congestion control mechanisms (Reno, CUBIC, Vegas, Tahoe, etc.).
- Authors assume that end-to-end route throughput is limited by the capacity of channels being load balanced, in other words those channels are considered a route bottle-neck. This assumption considerably limits the range of justifiable use cases. Route bottle-necks may often be several hops away and they may float in time because of congestions.

In [4] the author proposes an analytical solution for a method that allows minimizing the sum of delays of channels being aggregated. However, proposed solution is only valid for stationary case and doesn't take into account existing analytical models of different TCP congestion control mechanisms.

Our goal is to analyze efficiency of goodput aggregation of several channels of a multi-homed host using existing conventional Linux components for per-packet load balancing. The main emphasis is made on how delay and jitter of aggregated channels affect the efficiency of aggregation. We define aggregation efficiency as $E = \frac{A}{S}$, where A is the achieved goodput (bps) of several network channels aggregated using some technique; S is the sum of goodputs (bps) of those channels as used independently without any aggregation. Effects of aggregating channels with different throughputs and loss probabilities are analyzed in [7].

Note that per-flow load-balancing is traditionally preferred to per-packet load balancing, the latter considered undesirable because it can cause reordering of packets in a TCP session and may fail traversing NAT or firewall devices. However, per-packet load balancing allows finer-grained and more efficient load balancing allowing to achieve higher values of E .

2. Main section

To measure aggregation efficiency we use real life experiments instead of simulation because simulation results for TCP performance are non-probabilistic whereas real life TCP performance has high variance. In the hope of easy generalizations, we analyze aggregation of only two channels because adding more channels implies analysis of too many factors making real life experiments a burden because of their long lasting nature.

Experimental setup includes two virtual hosts: A and B, – both with guest OS “Linux Ubuntu 14.04”. Virtualizing software is “Oracle Virtual

Box 5.0.4” run on “MS Windows 7” (“Intel Core i7” processor with 8 GB of RAM). Host A has two Gigabit Ethernet adapters, eth0 and eth1, that are connected to eth0 (Gigabit Ethernet) of Host B via a virtual switch.

We use “iperf” utility to measure TCP performance with CUBIC congestion control mechanism in place (that’s the default in the above mentioned Linux distribution). We use “tc-netem” utility to set different throughput and RTT for eth0 and eth1 of host A, e.g. as follows:

```
— tc qdisc add dev eth0 root netem \
   rate 10000'kbit' delay 17'ms' 11'ms'
— tc qdisc add dev eth1 root netem \
   rate 5000'kbit' delay 10'ms'
```

The above commands limit eth0 capacity to 10 Mbps and on average add a 17 ms delay to each emitted packet varying that delay in the range from 6 to 28 ms. For eth1 the capacity is limited to 5 Mbps with 10 ms delay and no jitter. Note that mentioned delay is effectively a round-trip time (RTT) because no tc-netem configuration is done on host B.

All IP traffic coming from Host A to Host B is load-balanced between eth0 and eth1 using stochastic round robin rule. To do it we use “iptables” utility and its “mangle” table of the “OUTPUT” chain to label packets in *m:n* proportion using integer marks 1 and 2. For example, the following commands perform labeling in 4:1 proportion:

```
— iptables -t mangle -A OUTPUT \
   -m statistic --mode random --probability 1.0 \
   -j MARK --set-mark 1
— iptables -t mangle -A OUTPUT \
   -m statistic --mode random --probability 0.2 \
   -j MARK --set-mark 2
```

After that we use “iproute2” utility to route packets marked 1 and 2 to eth0 and eth1 respectively. The latter is achieved by using routing tables “RT_1” and “RT_2” manually added to file /etc/iproute2/rt_tables:

```
— ip route add table RT_1 default dev eth0
— ip rule add fwmark 1 table RT_1
— ip route add table RT_2 default dev eth1
— ip rule add fwmark 2 table RT_2
```

Since “iperf” measurements have high variation we have to increase experiment length as much as possible. However, increasing it makes experiments really long. Trying to make a trade-off between length of experiments and accuracy of results we set the length of one experiment to 200 s. In literature they recommend experiment length from 30 s (in [9]) up to 15 minutes (in [8]). Each our “iperf” experiment was used to get 10 samples from which we remove the lowest and the highest ones. The other 8 results are used to calculate confidence interval (using Student’s t-distribution with 95% confidence level) the width of which is put on figures in form of a vertical bars of corresponding length. If no vertical bars

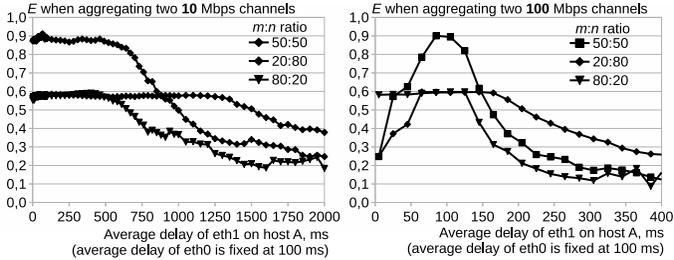


Figure 1. Goodput aggregation efficiency of channels with average delay asymmetry

are shown, it means we plot central moving average of respective values with window length 7 because confidence interval width is small enough.

On Fig. 1 are the results of experiments with equal-bandwidth channels in which we fixed average delay of one of them at 100 ms as average delay of the other was varied. We also varied balancing proportion $m:n$. It turned out that delay asymmetry indeed may degrade TCP performance if per-packet load balancing is used. As may be seen from Fig. 1, for certain delay asymmetry, E may become less than 0.5 thus making TCP performance worse than in the non-balancing scenario.

Another interesting point is that for slow channels (10 Mbps) there is an upper limit for delay below which E is guaranteed to be high even if delay asymmetry is 100-fold: see that E is close to 0.9 when delays in eth0 and eth1 are 1 and 100 ms respectively. At the same time for fast channels (100 Mbps) there is no such a limit: for small values of delay in eth1 we see that E degrades pretty fast.

Note also that for certain values of delay asymmetry, 50:50 balancing proportion is less optimal than 20:80. It means that even for equal-bandwidth channels one has to choose $m:n$ proportion carefully unlike authors of [5] and [6] propose.

On Fig. 2 are the results of experiments with equal-bandwidth channels in which we varied jitter of either only eth1 or both eth0 and eth1. We also varied balancing proportion $m:n$. It turned out that jitter asymmetry doesn't affect E at all. It seems that Linux TCP CUBIC has an efficient jitter buffer implementation which eliminates jitter asymmetry by increasing average delay. Thus jitter asymmetry while causing packet reordering doesn't degrade E . This may be considered as an unexpected result because TCP is known to have poor performance if packets arrive reordered.

On Fig. 3 are the results of experiments with equal-bandwidth channels (10 Mbps) in which we independently varied average delay of eth0 (d_0) and eth1 (d_1) for $m:n = 50:50$. Gray area corresponds to values of d_0 and d_1

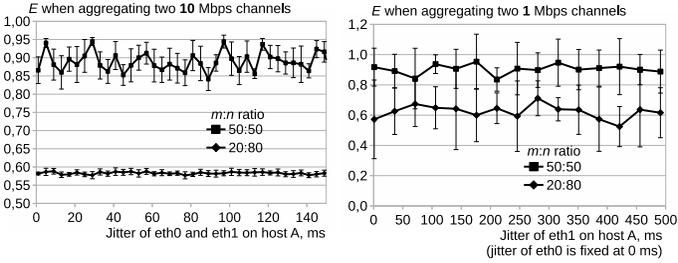


Figure 2. Goodput aggregation efficiency of channels with delay variation (jitter) asymmetry

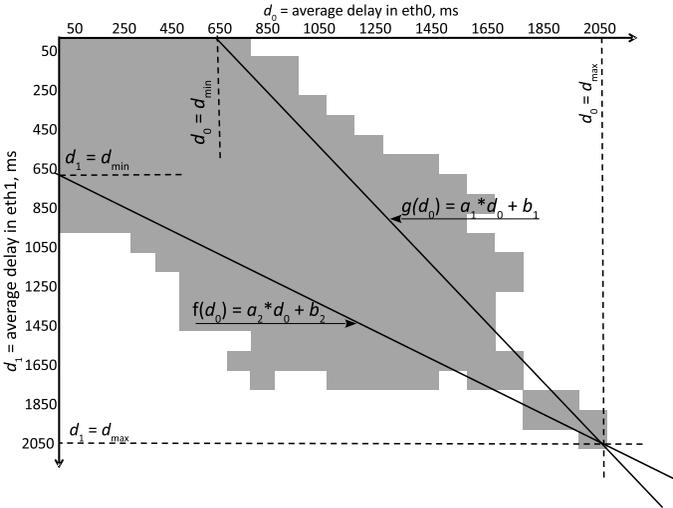


Figure 3. Goodput aggregation efficiency map for channels with average delay asymmetry

for which which measured value of E is more than 0.5. It means that in gray area one may actually increase TCP performance by using per-packet load-balancing in a multi-home environment. Considering symmetry of the experiment, we may safely affirm that gray area is symmetric relative to the $d_0 = d_1(d_0)$ line. Asymmetric distortions present on Fig. 3 are due to jittery nature of “iperf” measurements.

Experiments with other values of $m:n$ and capacities of channels may be summarized as follows:

- Gray area floats to the south-west if we decrease $m:n$. Gray area floats to the north-east if we increase $m:n$. It means we may compensate delay asymmetry by tuning the value of $m:n$.
- Gray area shrinks as capacity of channels increases. It means that per-packet load balancing of TCP is more effective on slower channels.

On Fig. 3 we also offer a rough approximation of gray area boundaries by two inequalities: $f(d_0) < d_1 < g(d_0)$, i.e. $a_1d_0 + b_1 < d_1 < a_2d_0 + b_2$. Unknown coefficients may be expressed via experimentally found values d_{max} and d_{min} as follows:

$$\left\{ \begin{array}{l} d_1 > \frac{d_0 d_{max}}{(d_{max} - d_{min})} + \frac{d_{max} d_{min}}{(d_{min} - d_{max})} \\ d_1 < \frac{d_0 (d_{max} - d_{min})}{d_{max} + d_{min}} \end{array} \right.$$

The values of d_{max} and d_{min} should be calculated individually because they depend on current maximum TCP window size allowed in a given operating system and on capacity of channels being aggregated.

The described results do not claim to be complete but nevertheless help to adumbrate the whole picture of goodput aggregation efficiency. To get a more clear picture one would need to take a huge amount of additional time-consuming experiments (e.g. Fig. 3 is based on a 17-hour experiment). A natural way to reduce experiment duration is to use simulation instead of real life experiments. Our plan is to continue research using a simulation model for verification of which we'll use real-life experimental data presented in this article.

3. Conclusions

As multihoming becomes a common practice, new technical solutions emerge to take advantage of environment in which there exist several network gateways. In this article we've experimentally evaluated goodput aggregation efficiency on a multihomed host using a conventional backward-compatible technique: per-packet load balancing of TCP traffic. While having inherent drawbacks (problems while traversing NAT servers or firewalls), this approach allows a fine-grained utilization scenario of network channels being aggregated. Experimental results showed that delay variation (jitter) asymmetry of aggregated channels doesn't affect the efficiency of goodput aggregation. On the contrary, average delay asymmetry turns out a significant factor of aggregation efficiency. We've presented an efficiency map of delay asymmetry and respective aggregation efficiency. The map implies that in certain conditions per-packet load-balancing is effective enough to be used instead of modern heavy-weight solutions (MPTCP, SCTP). Presented data may be used to verify simulation models of per-packet load balancing because no such data was available at the time of writing.

References

1. Network performance objectives for IP-based services. // International Telecommunication Union Recommendation ITU-T Y.1541. — Geneva, 2012. — 66 p.
2. *Raiciu C., etc.* Coupled Congestion Control for Multipath Transport Protocols // Internet Engineering Task Force Request for Comments. — No. 6356. — 2011. — 12 p.
3. *Adhari H., Dreibholz T., etc.* Evaluation of Concurrent Multipath Transfer over Dissimilar Paths // Proceedings of the 1st International Workshop on Protocols and Applications with Multi-Homing Support (PAMS). — 2011. — P. 708-714.
4. *Orekhov S.E., Sysoev I.V.* Optimizaciya raspredeleniya TCP-nagruzki po parallel'nomu setevomu traktu seti svyazi s mnogomernoj marshrutizaciej paketov // Izvestiya instituta inzhenernoj fiziki. — 2014. — No. 1 (31). — P. 57–59.
5. *Polyakov A.Y.* Adaptivnyj podhod k raspredeleniyu informacionnyh blokov po kanalam peredachi dannyh // EHlektrosvyaz' (ISSN 0013-5771). — 2009. — No. 6. — P. 32–35.
6. *Osipov A.V.* Distribution of Fragmented Frames over Redundant Channels of Computer Network // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. — 2013. — Vol. 13. — No. 2 (84). — P. 84–88.
7. *Sosnin V.V.* Analysis of TCP CUBIC Effectiveness for Capacity Aggregation in a Multi-Path Environment with Asymmetric Route Characteristics Using Per-Packet Load Balancing // Questions of Radio Electronics (ISSN 2218-5453). — No. 5. — 2016. — P. 24–31
8. Linux Foundation Networking Workgroup: TCP Testing. — URL: linuxfoundation.org/collaborate/workgroups/networking/tcp_testing (as visited on 2016.03.21).
9. *Constantine B., etc.* Framework for TCP Throughput Testing // Internet Engineering Task Force Request for Comments. — No. 6349. — 2011. — 27 p.

УДК 004.93

Метод оценки качества восприятия в системах дополненной реальности

В. А. Тельтевская, М. А. Маколкина

*Кафедра сетей связи и передачи данных,
Санкт-Петербургский университет телекоммуникаций им. проф. М.А.
Бонч-Бруевича,
пр. Большевиков, д.22, Санкт-Петербург, Россия, 193232*

Аннотация. В последнее время активное развитие получили технологии виртуальной и дополненной реальности. Для предоставления данной услуги необходимо передать данные, используемые в данных системах с приемлемым показателями качества обслуживания и качества восприятия. На данный момент не существует общепринятых методов оценки QoS и QoE в системах дополненной и виртуальной реальности и именно поэтому важно определить приемлемые показатели для QoS и QoE. В данной статье будет рассмотрен метод оценки качества восприятия в системах дополненной реальности с применением коэффициента Хёрста.

Ключевые слова: дополненная реальность, качество обслуживания, качество восприятия, коэффициент Хёрста, оценка качества.

1. Введение

В настоящее время, в связи с развитием инфокоммуникационных технологий, получили развитие технологии дополненной реальности. Дополненная реальность (англ. Augmented reality, AR) – это среда, созданная с дополнением физического мира виртуальными цифровыми сенсорными данными в режиме реального времени при помощи компьютерных устройств, а также программного обеспечения к ним. Она открывает для человека новые возможности восприятия и взаимодействия с окружающим миром, которые позволяют получить полезную информацию, применяя для этого минимум усилий. Предполагается, что применение данных технологий в современных инфокоммуникационных системах, откроет новый виток развития сетей связи пост-NGN и для услуг Интернета Вещей в частности [1, 2].

В свою очередь, для сетевых приложений, работающих в режиме реального времени, необходимо обеспечивать определенное качество обслуживания – QoS (англ. Quality of Service) [3]. Качество обслуживания определяется параметрами: полоса пропускания, задержка при передаче пакета, джиттер и потеря пакетов. Однако, пользователь услуги оценивает качество своим субъективным методом. Такая оценка называется оценкой качества восприятия (QoE). В системах дополненной реальности оценка пользователя выходит на первое место, т.к. её назначением является создание ощущения реального мира, который модернизирован для улучшения каких-либо характеристик

окружающего нас мира. Это может быть сделано с целью облегчения управления техническими или транспортными средствами, обучения, тренировки, игры, общения. Таким образом, в дополненной реальности должны оперативно обрабатываться данные, вовремя отображаться объекты, поступающие сигналы от датчиков должны передаваться без ошибок, движения объектов должны быть привычны человеческому глазу. В данной статье будет предпринята попытка установить взаимосвязь между качеством обслуживания и качеством восприятия.

Понятие качества является предметом стандартизации, в том числе качества услуг связи. [4–6]. Существуют стандарты и методы для оценки качества речи, видео, телевидения, но для направления дополненной реальности пока не было разработано ни одного стандарта. Следовательно, тема данной статьи является своевременной и актуальной.

Дополненная реальность – это новый перспективный виток в развитии инфокоммуникационных технологий. Поэтому, целью данного исследования стала разработка метода оценки качества восприятия в системах дополненной реальности. Для того, чтобы цель была достигнута, необходимо будет решить следующие задачи: изучить метод расчёта коэффициента Хёрста в сфере телекоммуникаций, создать систему дополненной реальности и определить качество восприятия для такой системы.

2. Разработка программного обеспечения

Для оценки качества восприятия в системах дополненной реальности было решено использовать коэффициент Хёрста [7, 8]. Коэффициент Хёрста – это мера, используемая в анализе временных рядов. Коэффициент будет уменьшаться в случае, если задержка между парой одинаковых значений временного ряда увеличивается. Применительно к теме данного исследования коэффициент отражает степень самоподобия трафика. Такой трафик обладает длительной памятью, т.е. если в течение некоторого времени в прошлом наблюдалось увеличение отклонения от среднего значения, то и в будущем будет происходить это увеличение. Иначе говоря, вероятность того, что трафик на следующем шаге отклонится от среднего значения в том же направлении, что и на предыдущем шаге, велика настолько, насколько коэффициент Хёрста приближен к единице. Итак, самоподобный трафик – это трафик, который выглядит одинаково при любых временных масштабах [10, 11].

Чтобы измерить коэффициент Хёрста для услуг дополненной реальности, необходимо перехватить и проанализировать трафик, генерируемый приложением дополненной реальности. Для этого, на базе модельной сети лаборатории Интернет Вещей Санкт-Петербургского

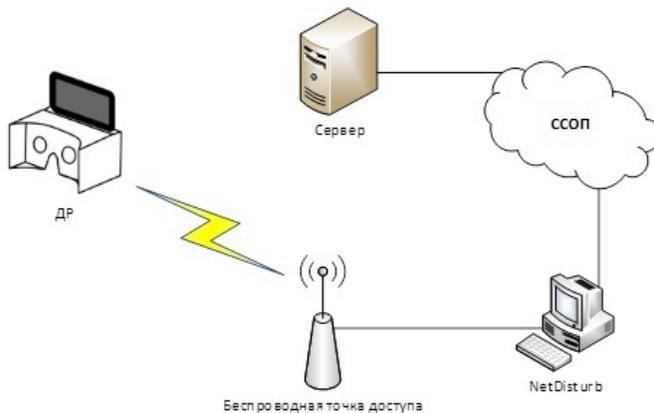


Рис. 1. Архитектура сети для эксперимента

государственного университета телекоммуникаций им. М.А.Бонч-Бруевича [9] было организовано соединение между сервером и устройством дополненной реальности (клиент). Для создания условий, максимально приближенных к реальному взаимодействию клиента и сервера в сети связи общего пользования (ССОП), в сеть был добавлен компьютер с возможностью создания искусственных помех с помощью ПО NetDisturb (рис. 1).

Для перехвата трафика, характерного для услуг дополненной реальности, было создано клиент-серверное приложение, которое позволяло бы клиенту наблюдать за перемещением виртуального объекта в режиме реального времени. Серверная часть этого приложения отвечает за получение, обработку и отправку координат местоположения объекта, при каждом обновлении кадра виртуальной реальности. После подключения клиента к серверу через беспроводную точку доступа с выходом в сеть Интернет, последний принимает запросы и отвечает на них, тем самым передавая координаты виртуального объекта. Приложение обрабатывает поступающие сообщения и перемещает объект в заданную точку. Благодаря постоянному взаимодействию клиента и сервера (запрос – ответ) появляется возможность установить корреляцию между QoS и QoE.

В свою очередь, в качестве клиента выступает устройство, используемое для визуального взаимодействия с пользователем. На нем хранится предварительно сконструированный виртуальный мир, включая объекты и их свойства. Данное устройство имеет гироскоп и акселерометр для системы наблюдения и управления дополненной реальностью.

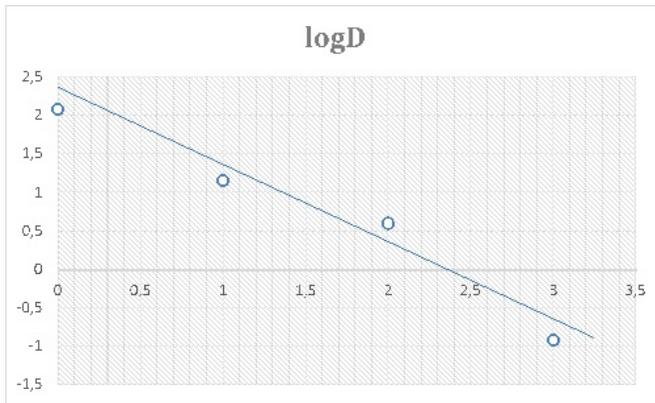


Рис. 2. Метод наименьших квадратов

Для оценки QoS и QoE достаточно двух характеристик перехваченных пакетов – время и размер. После того, как определены и записаны значения этих характеристик, можно приступить к их обработке. Для определения качества восприятия данной сети было решено использовать коэффициент Хёрста.

Для того, чтобы показатель Хёрста был точнее, перехват пакетов будет осуществляться в течение 1 минуты. Для временных рядов в 1 минуту будут высчитаны дисперсии (1)

$$\frac{\sum(C_i - \bar{C})^2}{n - 1} \quad (1)$$

– отклонение случайной величины от ее среднего значения в ряду. Чтобы точнее определить коэффициент Хёрста, интервал временного ряда будет изменен с 1 мс до 10, 100 и 1000 мс. Для каждого такого ряда рассчитывается дисперсия и с помощью метода наименьших квадратов (рис. 2) рассчитывается коэффициент a уравнения прямой. После этого можно рассчитать коэффициент Хёрста (2)

$$H = 1 + \frac{a}{2} \quad (2)$$

3. Результаты тестирования

С помощью созданного программного обеспечения в целях тестирования был проведён эксперимент по оценке качества восприятия в системе дополненной реальности. Был осуществлен перехват пакетов и их обработка при различных условиях работы сети – без помех, с

Таблица 1
Анализ эксперимента с добавлением задержки

Характеристика	Задержка				
	50мс	25мс	15мс	10мс	5мс
Частота обновления кадра	3	4	5	5	5
Скорость движения	1	3	4	4	5
Плавность движения	4	5	5	5	5
Полоса пропускания, кбит/с	6-8	11-13	15-17	18-22	23-26
Коэффициент Хёрста	0.463	0.523	0.561	0.602	0.610

задержками, с потерями, уменьшением пропускной способности. Во время проведения эксперимента за передвижением виртуального объекта в дополненной реальности следили 5 человек. Это было сделано с целью получить их субъективные оценки о плавности, скорости движения и частоты обновления кадра виртуального объекта. За эталон был принят эксперимент без внесения помех, т.е. все параметры в таком эксперименте имеют оценку 5. Все оценки пользователей были сведены в обобщающие таблицы со средней оценкой по каждому параметру:

1. Эксперимент с добавлением задержки (табл. 1).

Вывод: субъективная оценка пользователей показала, что задержки в 5 мс не оказывают негативного влияния на качество восприятия дополненной реальности. При задержке пакетов на 10-15 мс ухудшается восприятие скорости движения объекта, при больших задержках начинает уменьшаться частота обновления кадра. Для каждой величины задержки были измерены пропускная способность и коэффициент Хёрста. Из таб. 1 видно, что наибольший коэффициент Хёрста соответствует наивысшей оценке качества восприятия. Это означает, что трафик самоподобен и увеличения задержки, а, следовательно, и ухудшения качества восприятия, не произойдет с такой же вероятностью, насколько H приближен к 1. Для значений задержки больше 25 мс H принимает значение $< 0,5$. Такой трафик не является самоподобным и предсказать его поведение и оценку качества восприятия невозможно. Поэтому, в системах дополненной реальности допускаются задержки до 0,5 мс. Этот коэффициент можно использовать для оценки качества восприятия при задержках.

2. Эксперимент с потерями пакетов (табл. 2).

Вывод: потери пакетов негативно сказываются на оценке качества восприятия пользователей. Даже при 1% потерянных пакетов не достигается наивысшей субъективной оценки качества. Больше всего

Таблица 2

Анализ эксперимента с добавлением потерь

Потери				
Характеристика	10%	5%	2%	1%
Частота обновления кадра	1	1	2	3
Скорость движения	1	2	4	5
Плавность движения	1	1	1	2
Полоса пропускания, кбит/с	8-10	11-13	14-16	17-20
Коэффициент Хёрста	0.573	0.523	0.578	0.618

Таблица 3

Анализ эксперимента с понижением ПП

Пропускная способность				
Характеристика	5кбит/с	10кбит/с	20кбит/с	30кбит/с
Частота обновления кадра	2	4	5	5
Скорость движения	1	1	4	5
Плавность движения	5	5	5	5
Коэффициент Хёрста	0.480	0.525	0.574	0.587

страдают частота обновления кадров и плавность движения виртуального объекта. Это означает, что пользователь воспринимает изображение рывками, также пользователь наблюдает долгий отклик сцены на движения пользователя, например, поворот головы. Тем не менее, значения коэффициента Хёрста остается $> 0,5$. В связи с этим, напрашивается вывод, что коэффициент Хёрста нельзя использовать для связи QoS и QoE при наличии потерь пакетов. Однако, этот вывод может оказаться неверен, потому как если пакеты терялись на отрезке от сервера клиенту, а перехват трафика происходил с сервера, то значение H показывает, что пакеты отправлялись клиенту с сервера исправно.

3. Эксперимент с понижением пропускной способности (табл. 3).

Вывод: пропускная способность канала влияет на оценку качества восприятия дополненной реальности, и это отображает значение коэффициента Хёрста. Для созданной системы ДР достаточной пропускной способностью канала является 30 кбит/с. При снижении пропускной способности страдает скорость передвижения объекта. H отображает

зависимость оценки качества восприятия от оценки качества обслуживания при таком показателе качества, как пропускная способность.

4. Заключение

В рамках данной статьи был разработан метод для оценки качества восприятия в системах дополненной реальности. Для проведения исследования было также разработано специальное программное обеспечение для создания системы дополненной реальности и перехвата и анализа трафика.

При проведении тестирования и эксперимента с использованием созданного программного обеспечения для оценки качества восприятия в системах дополненной реальности, был сделан вывод, что такой метод подходит для оценки качества восприятия. Коэффициент Хёрста, рассчитанный по трафику, который отражает существующие недостатки сети, может быть использован для оценки качества восприятия в системах дополненной реальности. Также коэффициент Хёрста может быть использован для прогнозирования качества восприятия при планировании сетей.

Данный метод предполагается использовать для создания услуг, подразумевающих объединение дополненной реальности и Интернета Вещей.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ № 16-37-00209 в рамках научного проекта «Разработка принципов интеграции технологии дополненной реальности и Интернета Вещей».

Литература

1. *А.Е.Кучерявый, А.Е.Прокопьев, Е.А.Кучерявый* Самоорганизующиеся сети. – СПб.: Любавич, 2011. – 312 с.
2. *А.Е.Кучерявый* Интернет Вещей // Электросвязь. - 2013. - № 1. с.21-24.
3. Recommendation E.800 : Definitions of terms related to quality of service. ITU-T.
4. Recommendation P.800 : Methods For Subjective Determination Of Transmission Quality. ITU-T.
5. Recommendation P.910 : Subjective video quality assessment methods for multimedia applications. ITU-T.
6. ГОСТ Р 50691-2013 Услуги населению. Модель системы обеспечения качества услуг.
7. *Anastasia Vybornova, Andrey Koucheryavy* Traffic Analysis in Target Tracking Ubiquitous Sensor Networks // 14th International

- Conference, NEW2AN 2014 and 7th Conference, ruSMART 2014 vol. 8638 // Springer International Publishing // August 27–29, 2014.
8. *Парамонов А.И.* Разработка и исследование комплекса моделей трафика для сетей связи общего пользования // Автореферат, СПбГУТ // Санкт-Петербург 2014 год.
 9. *Kirichek R., Koucheryavy A.* Internet of Things Laboratory Test Bed // Lecture Notes in Electrical Engineering, 2016. Vol. 348. PP. 485–494.
 10. *Маколкина М. А., Прокопьев А. В.* Влияние коэффициента Хёрста на оценку качества восприятия видео // 69-я научно-техническая конференция Санкт-Петербургской общероссийской общественной организации «Российское научно-техническое общество радиотехники, электроники и связи» им. А. С. Попова (СПБНТОРЭС): тезисы докладов — СПб : СПБНТОРЭС, 2014. — С. 151–152.
 11. *Маколкина М. А., Прокопьев А. В., Кучерявый А. Е.* Параметр Хёрста как средство выявления аномальных изменений при передаче видео трафика // Актуальные проблемы инфотелекоммуникаций в науке и образовании. II-я международная научно-техническая и научно-методическая конференция: сб. научных статей / Под ред. Доценко С. М.. — СПб. : СПбГУТ, 2013. — С. 158–162.

UDC 004.93

Method for evaluating the quality of experience in augmented reality systems

V. A. Teltevszkaya, M. A. Makolkina

*The Bonch-Bruевич Saint-Petersburg State University of Telecommunications
Bolshevikov pr. 6, Saint-Petersburg, 193232, Russia*

Nowadays, there is the fast development of technology of augmented reality. To provide this service, it is necessary to provide not only the equipment, but also high quality data for rendering. Such data requires a special approach to the transmission, and therefore the high quality of service. It's needful to develop the new methods and standarts for this popular direction. In this article, there will be presented the method for evaluating the quality of experience in augmented reality systems using of the Hurst coefficient.

Keywords: augmented reality, quality of service, quality of experience, Hurst coefficient, evaluating the quality.

УДК 004.94

Гибридное имитационное моделирование модуля активного управления трафиком

Т. Р. Велиева*, А. В. Королькова*,
Д. С. Кулябов*†, Л. А. Севастьянов*‡

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

† *Лаборатория информационных технологий,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980*

‡ *Лаборатория теоретической физики,
Объединённый институт ядерных исследований,
ул. Жолио-Кюри 6, Дубна, Московская область, Россия, 141980*

Аннотация. Для исследования и верификации разрабатываемой коллективным математической модели сетевых протоколов были созданы дискретная имитационная и непрерывная аналитическая модели. Однако созданные реализации признаны не достаточно удовлетворительными. В связи с этим была поставлена задача разработать более адекватную имитационную модель, возможно в рамках иной парадигмы моделирования. В качестве подхода к моделированию сетевых протоколов предлагается использовать гибридный (непрерывно-дискретный) подход. Для компьютерной реализации модели использован язык физического моделирования Modelica. Открытый язык программирования Modelica имеет несколько реализаций. Нами выбран компилятор OpenModelica. Применение гибридного подхода позволяет в рамках непрерывной модели учесть переходы между разными состояниями при функционировании сетевых протоколов. Рассмотренный подход позволил получить простую имитационную модель протокола TCP. При этом данная модель имеет большой потенциал к расширению. Кроме того, представляется возможным использовать гибридный подход и при выполнении не только имитационного, но и аналитического моделирования.

Ключевые слова: гибридное моделирование, жидкостные модели, Modelica.

1. Введение

При моделировании сложных систем возникает проблема выбора модельного подхода. При применении дискретного и непрерывного подходов всегда остаются элементы модели, которые в не полной мере соответствуют выбранному подходу. Существующие модели систем с управлением не могут полностью удовлетворить нашим потребностям.

При моделировании сетевых протоколов перед нами встали серьёзные проблемы. Как оказалось, адекватные модели пакетного трафика просто отсутствуют. Нет даже общепринятой методики для его моделирования [1–3].

Для моделирования мы использовали непрерывные (жидкостные) модели TCP и RED [4, 5]. Однако, непрерывный подход позволяет лишь частично моделировать сетевые протоколы, так как обычно сетевые протоколы функционируют как конечные автоматы.

Для верификации полученной модели был использован программный комплекс ns-2 [6, 7], который позволяет проводить имитационное моделирование. По своей идеологии ns-2 является средством дискретного моделирования. В связи с этим комплекс обладает низкой масштабируемостью и подходит для моделирования малых сетей на малых промежутках времени.

Также проводилась верификация на натурном стенде на базе программного маршрутизатора [8].

Для дальнейшего развития нашей модели было принято решение использовать гибридный подход [9–13]. В данной статье описывается общий подход к гибридному моделированию сетевых протоколов. В качестве средства реализации предлагается использовать язык Modelica [14, 15].

В данной статье в разделе 2 описывается идеология гибридной парадигмы математического моделирования. В разделе 2.1 даётся общая информация об языке физического моделирования Modelica, реализующем в том числе и гибридную парадигму. В разделе 3 описывается процесс непрерывного и гибридного моделирования сетевого протокола TCP Reno.

2. Гибридный подход к моделированию

Гибридная¹ [9–13] система имеет как непрерывные, так и дискретные аспекты поведения. Гибридное поведение может быть обусловлено следующими причинами:

- совместное функционирование непрерывных и дискретных объектов. Например, система автоматического управления с непрерывным объектом управления и дискретным устройством управления.
- изменение состава системы. Здесь можно рассматривать системы с переменным количеством компонентов.
- мгновенные качественные изменения в непрерывном объекте. В данном случае качественные изменения непрерывных систем при моделировании представляются как дискретные события. В данном случае гибридность — не имманентная характеристика системы, а скорее приём моделирования.

Гибридные системы можно рассматривать как дискретно-непрерывные или как непрерывно-дискретные.

¹Другие названия: *непрерывно-дискретная система, система с переменной структурой, событийно-управляемая система.*

- В дискретных системах можно учитывать длительность ожидания очередного входного сигнала и длительность выходных действий (оснащение непрерывным временем).
- В модели с непрерывным временем можно допустить возможность протекания мгновенных процессов наряду с длительными.

Изначально нашим научным коллективом была разработана непрерывная динамическая модель. Для перехода к гибридной модели будем добавлять к существующей динамической модели дискретные элементы.

В непрерывных динамических моделях дискретные события могут задавать следующие составляющие:

- начальные условия или значения параметров в правых частях;
- форму правых частей;
- число уравнений.

Смена начальных условий и скачкообразное изменение параметров относятся к одному типу, поскольку скачкообразное изменение параметров может быть описано как смена начальных условий в некоторой новой системе уравнений.

Таким образом, в рамках гибридной модели можно представить как индикаторные функции, так и дифференциальные включения. Этот приём позволяет заменить систему с изменяющейся правой частью на систему с неизменной правой частью и меняющимися начальными условиями.

2.1. Язык моделирования Modelica

Язык Modelica [14, 15] позиционируется как объектно-ориентированный язык физического моделирования.

Modelica поддерживает непрерывную и гибридную [16] (непрерывно-дискретную) парадигмы. Впрочем, дискретные элементы в языке тоже присутствуют.

Язык Modelica представлен большим количеством коммерческих реализаций, такими как:

- Dymola [<http://www.claytex.com/>]. Один из наиболее развитых трансляторов для языка OpenModelica.
- CATIA [<http://www.3ds.com/products-services/catia/>]. Использует ядро на основе Dymola.
- MapleSim [<http://www.maplesoft.com/products/maplesim/>]. Интегрирована с Maple.
- Wolfram SystemModeler [<http://www.mathcore.com/>]. Является частью инфраструктуры платформы Wolfram.

Существуют и открытые реализации языка и среды Modelica:

- OpenModelica [<https://openmodelica.org/>]. Старейшая открытая реализация языка Modelica. Целью проекта является полная реализация спецификации Modelica.

- JModelica.org [<http://jmodelica.org/>]. Идеология реализации Jmodelica.org отличается от OpenModelica [17]. Целью проекта является возможность реализации расширений языка. В частности, Jmodelica.org поддерживает расширение Optimica, направленное на динамическую оптимизацию. Кроме того, связующим языком для Jmodelica.org является python.
- Modelicas реализует подмножество языка OpenModelica и используется в симуляторах Scicos [<http://www.scicos.org/>] и Xcos [<http://www.scilab.org/>].

3. Модель функционирования TCP

В протоколе TCP используется механизм скользящего окна для борьбы с перегрузками. Реализация данного механизма зависит от конкретного стандарта протокола TCP.

3.1. Механизм управления перегрузками в TCP

Поскольку оригинальная модель [8, 18, 19] базировалась на протоколе TCP Reno, то моделировать мы будем именно этот протокол.

В TCP Reno механизм управления перегрузками состоит из следующих фаз: медленный старт, предотвращение перегрузок, быстрая передача и быстрое восстановление. Динамика изменения размера окна перегрузки (Congestion Window, CWND) зависит от конкретной фазы.

Протокол TCP Reno отслеживает два варианта потери пакетов:

- Тройное дублирование подтверждения (Triple Duplicate ACK, TD).
- Тайм-аут (Timeout, TO).

Общий алгоритм управления перегрузкой относится к типу AIMD (Additive Increase, Multiplicative Decrease) — аддитивное увеличение размера окна и мультипликативное его уменьшение.

3.2. Переход к непрерывной модели для окна перегрузки

Поскольку мы строим гибридную непрерывно-дискретную модель, то для описания каждой фазы функционирования TCP перейдем к модели с непрерывным временем. Переход же между фазами будет описываться дискретными состояниями.

Используя сведения из раздела 3.1, формализуем поведение нашей модели.

Изменение размера TCP-окна связано с элементарным событием, которому соответствует либо приход одного подтверждения, либо приход всех подтверждений. Примем за элементарное событие приход всех подтверждений, который происходит за время двойного оборота (RTT).

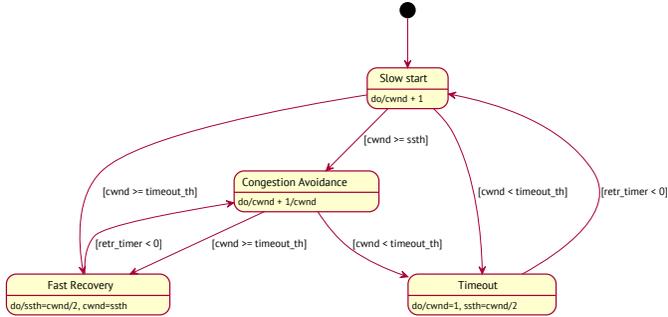


Рис. 1. Диаграмма состояний TCP

В фазе медленного старта (Slow Start) размер окна увеличивается при каждом приходе подтверждения АСК.

$$\frac{dW}{dt} = \frac{W}{T}, \quad d \ln W = \frac{dt}{T}, \quad \ln W = \frac{1}{T}; \quad W = \exp\left\{\frac{1}{T}\right\}.$$

Таким образом, окно растёт экспоненциально, как и должно быть в стадии медленного старта в соответствии с описанием TCP.

Аналогично рассмотрим фазу избежания перегрузок. При каждом приходе подтверждения АСК размер окна увеличивается.

$$\frac{dW}{dt} = \frac{1}{T}, \quad dW = \frac{dt}{T}, \quad W = \frac{1}{T}.$$

В результате получим линейный рост окна, как и описано в спецификации TCP.

3.3. Построение гибридной модели TCP

Для построения гибридной модели необходимо:

- записать динамическую модель для каждого состояния (сделано в пункте 3.2);
- заменить системы с кусочно-постоянными параметрами на системы с переменными начальными условиями;
- записать диаграмму состояний модели (рис. 1).

Полученную диаграмму (рис. 1) можно преобразовать в программу на языке Modelica (листинг 1).

Листинг 1: Алгоритм перехода состояний для протокола TCP

```
algorithm
state := TCPState.slowStart;
when edge(DelayD) and w >= timeout_th and (state ==
    TCPState.slowStart or state == TCPState.congestAvoid)
    then state := TCPState.fastRecov;
elsewhen w >= ssth and state == TCPState.slowStart then
    state := TCPState.congestAvoid;
elsewhen w < timeout_th and edge(DelayD) and (state ==
    TCPState.slowStart or state == TCPState.congestAvoid)
    then state := TCPState.timeOut;
elsewhen retr_timer < 0 and state == TCPState.fastRecov
    then state := TCPState.congestAvoid;
elsewhen retr_timer < 0 and state == TCPState.timeOut then
    state := TCPState.slowStart;
end when;
end TCPSender;
```

4. Заключение

В качестве реализации сетевого протокола нами рассматривался протокол TCP. Было сделано несколько математических моделей (как аналитических, так и имитационных) данного механизма с использованием разных парадигм и техник. При ближайшем рассмотрении во всех методах моделирования были выявлены определенные недостатки.

Рассмотренный в статье гибридный (непрерывно-дискретный) подход к моделированию представляется нам на данном этапе наиболее адекватным для моделирования сетевых протоколов.

Гибридный подход может применяться как при аналитическом, так и при имитационном моделировании. К сожалению, данный подход не достаточно активно применяется исследователями, хотя и реализован в ряде систем компьютерного моделирования.

Благодарности

Работа частично поддержана грантами РФФИ № 14-01-00628, 15-07-08795 и 16-07-00556.

Литература

1. Paxson V., Floyd S. Why we don't know how to simulate the Internet // Proceedings of the 29th conference on Winter simulation - WSC '97. — New York, New York, USA : ACM Press, 1997. — P. 1037–1044.

2. Paxson V., Floyd S. Wide area traffic: the failure of Poisson modeling // IEEE/ACM Transactions on Networking. — 1995. — jun. — Vol. 3, no. 3. — P. 226–244.
3. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V. On the self-similar nature of Ethernet traffic (extended version) // IEEE/ACM Transactions on Networking. — 1994. — Vol. 2, no. 1. — P. 1–15.
4. Demidova A. V., Korolkova A. V., Kulyabov D. S., Sevastyanov L. A. The method of constructing models of peer to peer protocols // 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). — IEEE, 2014. — P. 557–562. — 1504.00576.
5. Eferina E. G., Korolkova A. V., Gevorkyan M. N. et al. One-Step Stochastic Processes Simulation Software Package // Bulletin of Peoples' Friendship University of Russia. Series "Mathematics. Information Sciences. Physics". — 2014. — no. 3. — P. 46–59. — 1503.07342.
6. Altman E., Jiménez T. NS Simulator for Beginners // Synthesis Lectures on Communication Networks. — 2012. — Vol. 5, no. 1. — P. 1–184.
7. Issariyakul T., Hossain E. Introduction to network simulator NS2. — 2012. — Vol. 9781461414. — P. 1–510.
8. Velieva T. R., Korolkova A. V., Kulyabov D. S. Designing installations for verification of the model of active queue management discipline RED in the GNS3 // 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). — IEEE, 2014. — P. 570–577. — 1504.02324.
9. Maler O. Hybrid Systems and Real-World Computations // Workshop on Theory of Hybrid Systems. — Lyndby, Denmark : Springer-Verlag, 1992.
10. Maler O. Control from computer science // Annual Reviews in Control. — 2002. — jan. — Vol. 26, no. 2. — P. 175–187.
11. Färnqvist D., Strandemar K., Johansson K. H., Hespanha J. P. Hybrid Modeling of Communication Networks Using Modelica // The 2nd International Modelica Conference. — 2002. — P. 209–213.
12. Hespanha J. P., Bohacek S., Obraczka K., Lee J. Hybrid Modeling of TCP Congestion Control // Lncs. — 2001. — No. 2034. — P. 291–304.
13. Bohacek S., Lee J. Analysis of a TCP hybrid model // Proc. of the 39th Annual Allerton Conference on Communication, Control, and Computing. — 2001. — P. 1–10.
14. Fritzson P. Principles of Object-Oriented Modeling and Simulation with Modelica 2.1. — Wiley-IEEE Press, 2003. — P. 939.
15. Fritzson P. Introduction to Modeling and Simulation of Technical and Physical Systems with Modelica. — Hoboken, NJ, USA : John Wiley & Sons, Inc., 2011.
16. Broman D., Greenberg L., Lee E. A. et al. Requirements for hybrid cosimulation standards // Proceedings of the 18th International Conference on Hybrid Systems Computation and Control - HSCC '15. —

2015. — no. April. — P. 179–188.
17. Broman D., Fritzson P., Hedin G., Åkesson J. A comparison of two metacompilation approaches to implementing a complex domain-specific language // Proceedings of the 27th Annual ACM Symposium on Applied Computing - SAC '12. — New York, New York, USA : ACM Press, 2012. — P. 1919.
 18. Misra V., Gong W.-B., Towsley D. Stochastic differential equation modeling and analysis of TCP-window size behavior // Proceedings of PERFORMANCE. — 1999. — Vol. 99.
 19. Misra V., Gong W.-B., Towsley D. Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED // ACM SIGCOMM Computer Communication Review. — 2000. — Vol. 30, no. 4. — P. 151–160.

UDC 004.94

Hybrid simulation of active traffic management

T. R. Velieva*, A. V. Korolkova*, D. S. Kulyabov*[†],
L. A. Sevastianov*[‡]

* *Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

[†] *Laboratory of Information Technologies
Joint Institute for Nuclear Research
Joliot-Curie 6, Dubna, Moscow region, 141980, Russia*

[‡] *Bogoliubov Laboratory of Theoretical Physics
Joint Institute for Nuclear Research
Joliot-Curie 6, Dubna, Moscow region, 141980, Russia*

For the study and verification of our mathematical model of network protocols a discrete simulation model and a continuous analytical model were developed. However, for various reasons, these implementations are not entirely satisfactory. In connection with the task to develop a more adequate simulation model, possibly using a different modeling paradigm. In order to modeling of networks protocols it is proposed to use a hybrid (continuous-discrete) approach. For computer implementation of the model the physical modeling language Modelica is used. Because the language Modelica has multiple implementations we have selected the OpenModelica compiler. The hybrid approach allows us to take into account the transitions between different states in the continuous model of network protocols. The considered approach allowed to obtain a simple simulation model of the TCP network protocol. This model has great potential for expansion. Furthermore, it is possible to use a hybrid approach not only for the simulation but also for analytical modeling.

Keywords: hybrid modeling, fluid model, Modelica.

УДК 004.722.25

Оптимизация расписания передачи данных в опорных сетях 5G миллиметрового диапазона под управлением схемы STDMA

В. М. Вишневский, Р. Е. Иванов,
А. А. Ларионов, М. С. Дудин

*Институт Проблем Управления РАН им. Трапезникова,
ул. Профсоюзная, д.65, Москва, Россия, 117997*

Аннотация. В статье рассматриваются варианты построения опорной сети 5G на базе mesh-сети миллиметрового E-диапазона радиоволн. Описывается возможность использования технологии ММО, обеспечивающей формирование сразу нескольких разнонаправленных лучей, а также возможность частотного дуплексирования, позволяющего каждой станции mesh-сети одновременно осуществлять и приём, и передачу. Рассматривается структура суперкадра и схема STDMA доступа к среде для разрабатываемой mesh-сети. Описывается решение задачи распределения частотно-временных ресурсов в опорных сетях миллиметрового E-диапазона с использованием графового подхода и приводится эвристический алгоритм повышения эффективности этого распределения.

Ключевые слова: 5G, опорная сеть, mesh-сеть, мм-диапазон радиоволн, FDD, TDMA, теория графов.

1. Введение

Широкополосные беспроводные сети нового, пятого поколения (5G), внедрение которых ожидается к 2020 г., стали в настоящее время одним из основных направлений развития телекоммуникационной индустрии. Основным отличием сетей этого класса от беспроводных сетей предыдущих поколений является многократное повышение пропускной способности в связи с лавинообразным ростом передаваемой информации, возможность одновременного обслуживания большого количества соединений, связанное с появлением таких новых технологий, как Интернет вещей (IoT) и интеллектуальные транспортные системы (VANET), повышение качества и предоставление целого ряда новых услуг, среди которых виртуальная и дополненная реальность (Virtual/Augmented Reality).

В рамках создания сетей 5G особое внимание уделяется возможности использования миллиметрового диапазона радиоволн, в частности, E-диапазона (71–76 ГГц, 81–86 ГГц). В отличие от дефицитного сантиметрового диапазона, в котором функционируют существующие

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 16-57-00130).

широкополосные беспроводные сети, с жестким ограничением на ширину частотной полосы, миллиметровый диапазон обеспечивает возможность выделения широких полос частот (до 10 ГГц) и, соответственно, сверхвысокую скорость беспроводной передачи информации (до 50 Гбит/с).

Другими преимуществами систем передачи информации в этом диапазоне являются:

- создания многоэлементных антенных систем;
- свойственное данному диапазону быстрое затухание радиоволн и узкая направленность антенных систем позволяет увеличить спектральную эффективность за счёт пространственного переиспользования частот (значительно упрощает решение задачи частотного планирования);
- возможность создания систем связи, обеспечивающих скрытность связи (необнаруживаемость) и целостность (стойкость к прицельным помехам и несанкционированным подключениям).

В связи с описанными преимуществами в последние годы появилось значительное количество работ, посвященных исследованию вариантов построения сетей 5G в миллиметровом диапазоне радиоволн [1–3]. В настоящей статье, являющейся развитием работ авторов [4], рассматриваются варианты построения опорной сети 5G на базе mesh-сети миллиметрового E-диапазона радиоволн. Описывается возможность использования технологии MIMO, а также возможность частотного дуплексирования, позволяющего каждой станции mesh-сети одновременно осуществлять и прием и передачу. Рассматривается структура суперкадра и схема TDMA доступа к среде для разрабатываемой mesh-сети. Поставлена и решена задача распределения частотных ресурсов в опорных сетях E-диапазона радиоволн, сформированная в терминах теории графов. Приводится эвристический алгоритм решения этой задачи, обеспечивающий повышение эффективности распределения ресурсов сети.

2. Принципы построения опорных сетей в E-диапазоне радиоволн

Любая технология связи, используемая в качестве опорной сети в сотовых сетях пятого поколения, обязана предоставлять сотовой сети сервис передачи данных, совместимый с требованиями, предъявляемые к сотовой сети в целом. На данный момент организация ITU IMT-2020 (5G) Promotion Group выделила ряд требований к сетям 5G [5], среди которых:

- спектральная эффективность: 45 bps/Hz
- плотность трафика: 100 Tbps на квадратный километр
- число подключений: 1 миллион устройств на километр
- минимальная скорость на абонента: 100 Mbps
- задержка: меньше 1 мс

Схожие требования предлагает также один из основныз разработчиков сотовых сетей 5G - Samsung Electronics Co. [6]

Помимо обслуживания традиционных полнофункциональных базовых станций сотовой сети (в терминологии LTE называемых evolved Node B или eNB), формирующих распределенную сеть радио-доступа (Distributed Radio Access Network, D-RAN) и подключаемых к ядру операторской сети по соединениям, называемым backhaul, опорная сеть должна обеспечивать создание централизованных сетей радио-доступа (Centralized Radio Access Network, C-RAN) [1]. В таких сетях часть функций обработки сигналов переносится на устройства BBU (Base Band Unit), к каждому из которых подключается несколько более дешевых устройств RRU (Remote Radio Unit), которые используются для непосредственного подключения абонентов. Соединения между RRU и BBU называются fronthaul (см. рис. 1). Эти соединения, с одной стороны, в среднем короче, чем backhaul, однако имеют более жесткие требования к задержке и ее вариации (джиттеру).

Из приведенных требований следует, что опорная сеть обязана предоставлять пропускную способность не менее 1 Gbps каждой станции сотовой сети, обслуживающей не более 10 абонентов, и пропускную способность свыше 5-10 Gbps для традиционных базовых станций, поддерживающих подключение большего числа абонентов. Также необходимо учитывать высокое число малых базовых станций сотовой сети, без которых будет невозможно обеспечить высокую плотность трафика и подключений, которые будет необходимо подключить с помощью опорной сети к ядру сотовой сети. Эти требования и особенности делают миллиметровый E-диапазон наиболее целесообразным решением для создания опорной широкополосной беспроводной сети, обеспечивающей подключение базовых станций сотовой сети 5G.

Жесткие требования к задержкам передачи данных, в особенности обслуживании fronthaul-соединений, затрудняют использование стандартных схем доступа к среде на основе CSMA/CA, широко используемых в сетях сантиметрового диапазона. Для обеспечения низких задержек целесообразно использовать схемы доступа, основанные на планировании доступа к радиосреде, в которых значительно снижается вероятность коллизий и, как следствие, потери пропускной способности и увеличения задержек и их вариаций из-за повторных передач кадров. В качестве схемы доступа можно использовать комбинацию Spatial TDMA (STDMA) для разделения времени между передачами соседним станциям, и схему частотного дуплексирования FDD для создания симметричных каналов связи. Помимо минимизации числа коллизий, STDMA позволяет организовать передачи данных по многоточечным соединениям, поочередно формируя луч для связи с соседними станциями опорной сети.

В настоящей работе предлагается использовать в качестве опорной сети для подключения базовых станций сотовой сети пятого поколения самоорганизующуюся MESH-сеть миллиметрового E-диапазона

радиоволн [4]. Такая сеть формируется из равноправных станций, которые самостоятельно подключаются к сети, сканируют эфир, создают и поддерживают соединения с другими станциями сети. Соединения в mesh-сети бывают как одношаговые, так и многошаговые, построение и управление которыми осуществляется с помощью протоколов маршрутизации и распределения нагрузки. Mesh-сети обладают высокой надежностью за счет резервирования соединений, имеют низкую стоимость и высокую скорость развертывания. В качестве схемы доступа в предлагаемой сети используется Spatial TDMA и частотное дуплексирование, а передача данных станциями осуществляется погласно расписаниям. Следует отметить, что базовые станции сотовой сети могут подключаться к станциям mesh-сети как с использованием проводных/беспроводных интерфейсов, так и быть реализованными с ними в одном корпусе.

Использование частотного дуплексирования FDD позволяет увеличить пропускную способность за счет формирования симметричных каналов связи между соседними станциями. При этом каждая станция mesh-сети выбирает, в какой полосе частот ей вести передачу, а в какой - принимать. В зависимости от сделанного выбора станция попадает в одну из двух групп. Станции из одной группы, использующие одни и те же полосы для передачи (соответственно, и для приема), не могут формировать соединения. Первоначальный выбор частот производится станцией при подключении к сети, однако в дальнейшем он может быть изменен. При переназначении частоты хотя бы одной станции, топология сети изменяется. Это свойство будет использовано в дальнейшем для оптимизации распределения частотно-временных ресурсов сети.

В качестве схемы множественного доступа к среде выбрана STDMA, так как она обеспечивает минимизацию числа коллизий и, в то же время, соответствует принципу работы антенных систем: радиосигнал в E-диапазоне является узконаправленным и для подключения к соседней станции антенне необходимо отклонять сформированный луч в нужном направлении. При этом вся полусфера вокруг станции делится на сектора, ширина и высота которых равны ширине и высоте луча соответственно. Антенные массивы, состоящие из небольшого числа элементов, способны излучать сигнал только в один сектор одновременно и если станции находятся в разных секторах, то связаться с ними можно только в различные интервалы времени.

Следует отметить, что при использовании более дорогих антенн технологии massive MIMO, станция приобретает возможность формировать сразу несколько разнонаправленных лучей, позволяя частично или полностью преодолеть описанное выше ограничение. Узконаправленность сигнала и быстрое затухание позволяют в полной мере реализовать пространственное переиспользование полосы частот, позволяя разным станциям передавать сообщение на одной частоте и в одно время.

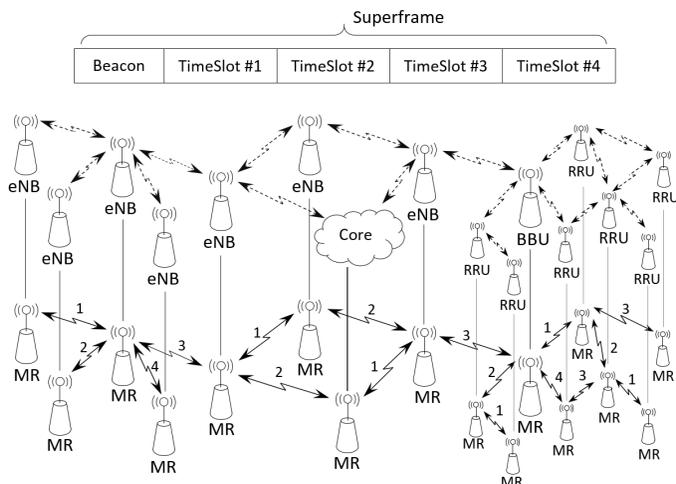


Рис. 1. График зависимости среднего числа станций, которые необходимо перекастить, от числа станций в сети

Согласно схеме STDMA, все время работы сети разбивается на суперкадры (см. рис. 1). Каждый суперкадр начинается со специального beacon-интервала, предназначенного для сканирования эфира и формирования соединений, за которым идет последовательность слотов. Каждой станции для связи с соседями назначаются номера слотов. В каждом слоте любая станция ведет передачу не более, чем одному соседу. Назначение номеров слотов соединениям называется расписанием. Если топология сети не изменяется, расписание повторяется в последовательных суперкадрах.

Использование расписаний для разделения передачи данных по разным соединениям позволяет значительно снизить риск коллизий, сократить вариацию задержки, а также гарантировать выделение заданной пропускной способности соединениям. В то же время, в этой схеме возможен простой станций в том случае, если число соединений от станции к станции сети сильно различается: если каждому соединению выделяется по одному слоту в расписании, то число слотов определяется станциями, имеющими наибольшее число соединений. При этом станции, у которых соседей меньше, будут вести передачу в меньшем числе слотов, а в неиспользованных слотах эти станции будут простаивать, из-за чего будет происходить потеря общей пропускной способности. Существует несколько подходов к решению этой проблемы:

1. распределение свободных слотов между имеющимися соединениями

2. изменение топологии сети для сокращения вариации числа соединений

Первый подход может быть просто реализован, однако может вести к образованию “бутылочного горла”, если не использовать его совместно с механизмами балансировки нагрузки. В настоящей статье предлагается использовать второй подход – метод изменения топологии сети с помощью смены частот на отдельных станциях для минимизации числа “пустующих” слотов. Для формальной постановки задачи, описания и анализа алгоритма требуется ввести формальную математическую модель сети. В следующем разделе задача будет сформулирована с использованием графового подхода.

3. Графовая модель опорной mesh-сети

Пусть заданы местоположения всех станций mesh-сети, и для каждой пары станций известно, будет ли между ними существовать соединение, если этим станциям назначить соответствующие частоты приема/передачи. Построим граф $G = (V, E)$, который будет использоваться для получения реальных топологий mesh-сети. Для этого поставим во взаимнооднозначное соответствие совокупности станций множество вершин V , а множество ребер E образуем из тех и только тех пар вершин (u, v) , между прообразами которых (то есть станциями mesh-сети) будет существовать соединение при соответствующем назначении станциям частот, и это соединение будет отвечать заданным при настройке сети параметрам, таким как пропускная способность, надежность, задержка. Любую допустимую топологию mesh-сети на заданных станциях можно описать в виде остоного подграфа графа $G = (V, E)$.

В предлагаемой архитектуре сети используется частотное дуплексирование с двумя полосами частот: в зависимости от того, какая полоса используется для приёма, а какая – для передачи, станцию можно отнести к одной из двух групп, при этом соединения могут существовать только между станциями разных типов. Так как в графе такой сети ребра будут соединять вершины из разных групп, такой граф по определению является двудольным. Таким образом, произвольная топология mesh-сети на заданных станциях описывается как остовой двудольный подграф $B = (V, E')$, $E' \subset E$ графа G .

Поскольку конкретная топология mesh-сети определяется назначением частот станциям, она будет зависеть от метрик, используемых при выборе частот станцией, подключаемой к сети. Например, новая станция может выбрать ту пару частот, которая позволит ей организовать наибольшее число подключений к соседям (жадный метод повышения связности), выбирать частоты исходя из качества соединений или каким-то иным способом. Поскольку описание как самого процесса подключения станции, так и используемых в нем метрик выходят за рамки настоящей статьи, будем считать, что подключаемая станция определяет частоты случайным равновероятным образом. Это

позволяет считать все остовные двудольные подграфы B графа G равновероятными.

Построим расписание сети, то есть назначим каждому соединению интервал времени, в который она может передать сообщение. Для этого поделим период работы сети – суперкадр – на T временных слотов (структура отображена на рис. 1). Для построения расписания необходимо назначить каждому соединению свой номер слота, в котором станция сможет передать данные соседу. Данную задачу можно сформулировать как задачу рёберной раскраски графа. Для этого определим в качестве цвета ребра номер временного слота. Хотя в общем случае задача реберной раскраски графа NP-полна, для двудольного графа она оказывается полиномиальной: согласно теореме Визинга, $T = D = D(B) = \max_{v \in V} \deg(v)$. Для построения раскраски существует ряд полиномиальных алгоритмов, например – алгоритм Гэбоу [7, 8].

Во время работы сети состав станций и связи между ними могут меняться. При этом если возникает необходимость выбора частотного режима, станция старается максимизировать число связей с соседями. К сожалению, такая схема приводит к необходимости оптимизировать все расписание. В следующей главе поставлена и решена задача максимизации пропускной способности.

4. Задача максимизации пропускной способности mesh-сети

Как было показано в конце раздела 2, из-за различного числа соседей некоторые станции не смогут использовать все время работы для передачи данных: станции, у которых меньше соединений, будут использовать лишь те слоты, которые им назначены, а в остальное время (когда станции с большим числом соседей будут продолжать вести передачу) будут простаивать. Эта потеря пропускной способности происходит на тех станциях, число соединений которых меньше числа слотов. В терминах двудольного графа сети B , потеря происходит на вершине $v \in V$, для которой $d = \deg(v) < D$, и составляет $D - \deg(v)$. Таким образом, суммарное число потерянных слотов – функция от графа B , которую можно определить как

$$M(B) = \sum_{v \in V} D(B) - \deg(v) = N(D(B) - d_{mean}),$$

где N – число станций в сети, в d_{mean} – средняя степень вершины. Таким образом, потеря пропускной способности будет наименьшей на том подграфе B , для которого $\mu(B) = D(B) - d_{mean} \rightarrow \min$, наиболее близкому к регулярному графу.

Пусть \mathfrak{B}_G – множество всех связных остовных двудольных подграфов графа G . Задача поиска оптимального с точки зрения максимизации пропускной способности распределения частотных ресурсов в

mesh-сети может быть сформулирована как задача поиска двудольного остовного подграфа B^* , для которого:

$$B^* = \arg \min_{B \in \mathfrak{B}_G} \mu(B)$$

Очевидно, что оптимальным с точки зрения приведенного критерия будет двудольный связный k -регулярный подграф, если он существует. К сожалению, уже ответ на вопрос о существовании такого подграфа является NP-полной проблемой [9]. Поэтому для поиска оптимальной топологии приходится использовать приближенные методы.

Вместо поиска оптимального графа B^* , поставим задачу улучшить пропускную способность существующей сети, описываемой графом B_0 . Для этого будем изменять частоты на некотором подмножестве станций. Предположим, что в сети изменяются назначения частот k станций. Измененной сети соответствует граф B_0^k , отличающийся от B_0 тем, что k вершин в нем находятся в другой доле. В случае, если $\mu(B_0) < \mu(B_0^k)$, сделанное изменение частот ведет к увеличению пропускной способности сети. Пусть $x(B) \in \{0, 1\}^N$ – вектор, поставленный в соответствие двудольному подграфу B : $x_i(B) = x_j(B)$ тогда и только тогда, когда вершины i, j находятся в одной доле. Пусть $\rho(x(B_1), x(B_2))$ – расстояние Хемминга между векторами, кодирующими графы B_1, B_2 . Используя введенные обозначения, можно утверждать, что $\rho(x(B_0), x(B_0^k)) = k$, то есть вектор, кодирующий граф сети, полученной изменением частот на k станциях, находится на расстоянии Хемминга от исходного графа, равном k .

Поскольку, как было отмечено в разделе 3, все двудольные связные графы $B \in \mathfrak{B}_G$ считаются равновероятными, расстояние Хемминга до ближайшего оптимального графа B^* (оптимум может достигаться на нескольких подграфах) будет различным для различных подграфов. Вместе с тем, можно говорить о среднем по всему множеству \mathfrak{B}_G расстоянии до оптимального подграфа, которое, очевидно, должно зависеть от числа станций сети N . Если K – это среднее расстояние, то разумно ограничиться изменением частот на не более, чем K станциях сети – в среднем, это приведет к оптимизации пропускной способности.

Поскольку $|\{B' \in \mathfrak{B}_G : \rho(B', B_0) \leq K\}| \sim O(N^K)$, целесообразно сократить количество станций, на которых изменяются частоты. Для этого помимо поиска среднего расстояния K до оптимальных графов, будем искать расстояния $K_\delta = 1/N \sum_{B \in \mathfrak{B}_G} \rho(B, B_\delta^*)$ до графов B_δ^* , для которых $\mu(B_\delta^*) - \mu(B^*) \leq \delta(\mu(B_0) - \mu(B^*))$, то есть среднее число станций, частоты на которых нужно сменить для того, чтобы уменьшить отличие от оптимума в δ^{-1} раз.

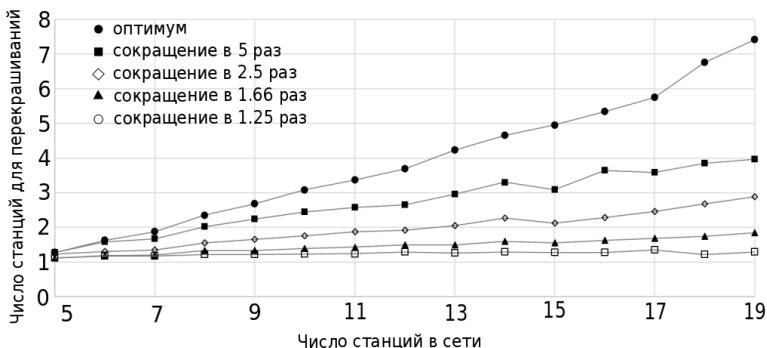


Рис. 2. Архитектура сети backhaul на основе mesh-сети мм-диапазона

5. Численное исследование

Для численного анализа предложенного метода была разработана программа на языке Python 3, которая позволяет генерировать произвольные mesh-сети с заданным числом вершин, вычислять значения метрики регулярности $\mu(B)$, находить оптимальный подграф B^* и находить число станций, на которых необходимо изменить частоты, для получения оптимального графа B^* , либо графа B_δ^* . Результаты численного исследования приведены на рис. 2: показана зависимость оптимального числа перекрашиваний K от числа базовых станций в сети, а также аналогичные зависимости K_δ для различных значений дельта, позволяющих добиться 5-, 2.5-, 1.66 и 1.25-кратного сокращения расстояния между данной и оптимальной топологиями. Как видно из рисунка, данные зависимости близки к линейным и при этом для того, чтобы прийти к оптимальной топологии, в среднем требуется в два раза больше перекрашиваний, чем для того, чтобы сократить расстояние до оптимума в 5 раз для сети, состоящей из 19 станций. Поиск оптимальной топологии может потребовать много времени, но можно добиться значительного увеличения производительности сети, ограничившись лишь 3 или 4 сменами частотного плана.

6. Заключение

В статье рассмотрены варианты построения опорной mesh-сети миллиметрового E-диапазона радиоволн. Построена графовая модель, позволяющая оптимизировать частотно-временные характеристики, а также топологию опорной сети. Предложен эвристический подход к отысканию квазиоптимальных решений. Приведены численные результаты исследования предложенного эвристического подхода.

Литература

1. M. Jaber, M.A. Imran, R. Tafazolli, A. Tukmanov, 5G Backhaul Challenges and Emerging Research Directions: a Survey. IEEE Access, 2016, Vol. 4, pp. 1743-1766.
2. Yong Niu, Yong Li, Depeng Jin, Li Su, Athanasios V.Vasikos, A Survey of Millimeter Wave (mmWave) Communication for 5G: Opportunities and Challenges. Wireless Networks, 2015, Vol. 21, Is 8, pp. 2657-2676.
3. Kan Zheng, Long Zhao, Jie Mei, Mischa Dohler, Wei Xiang, Yuexing Peng, 10 Gb/s HetSNETs with Millimeter-Wave Communications: Access and Networking – Challenges and Protocols. IEEE Communication Magazine, 2015, Vol. 53, Is. 1, pp. 222-231.
4. V. Vishnevsky A. Larionov, S. Frolov, Design and Scheduling in 5G Stationary and Mobile Communication System Based on Wireless Millimeter-Wave Mesh Networks. Communications in Computer And Information Science. 2014, Vol. 279, pp. 11-28.
5. 5G Vision of IMT-2020 (5G) Promotion Group. www.imt-2020.cn/en/
6. Samsung 5G Vision. www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-2.pdf
7. H.N. Gabow, Using Eulerian partitions to edge color bipartite graphs. Internat. J. Comput. Inform. Sci., 5 (1976), pp. 345–355.
8. H.N. Gabow, O. Kariv, Algorithms for edge coloring bipartite graphs and multigraphs. SIAM J. Comput., 11 (1982), pp. 117–129.
9. F. Cheah, D.G. Corneil, The complexity of regular subgraph recognition. Discrete Applied Mathematics, May 1990. Vol. 27, Iss. 1–2, pp 59-68.

UDC 004.722.25

Optimisation of data transmission scheduling in 5G mmWave backhaul network with STDMA

V. M. Vishnevsky, R. E. Ivanov, A. A. Larionov, M. S. Dudin

*V.A. Trapeznikov Institute of Control Sciences RAS,
Profsoyusnaya str. 65, Moscow, 117997, Russia*

Different variants of E-band mmWave backhaul design are presented in this paper. The way of making use of MIMO technology that provides multi-beam transmission are described as well as the usage of FDD allowing simultaneous transmission and reception for each station of a mesh-network. The structure of superframe and STDMA as media access scheme for developing mesh-network are concerned. The problem of time-frequency resource scheduling in E-band mmWave backhaul network are described by applying theoretic-graph approach and the heuristic algorithm of the schedule performance enhancement are presented.

Keywords: 5G, backhaul, mesh, mmWave, FDD, TDMA, graph theory.

УДК 519.218.31

Интервальные оценки характеристик системы с оптимальным выбором

Я. Г. Серж*, Т. А. Милованова*, И. С. Зарядов*†

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклуто-Маклая, д.6, Москва, Россия, 117198*

† *Институт проблем информатики ФИЦ ИУ РАН,
ул. Вавилова 44 кор.2, ИПИ ФИЦ ИУ РАН, Москва, Россия, 119333*

Аннотация. Исследуется система массового обслуживания, состоящая из двух параллельных подсистем (каждая подсистема включает в себя накопитель и обслуживающий прибор). Дисциплина оптимального выбора заключается в следующем: поступающие в систему заявки выбирают ту подсистему, для которой длина очереди в накопителе минимальна. Если в обеих подсистемах длины очередей совпадают, то с некоторой вероятностью выбирается одна из подсистем, с дополнительной вероятностью выбирается вторая подсистема. Предполагается, что в систему поступает пуассоновский поток, времена обслуживания на приборах подчинены экспоненциальному распределению с различными значениями параметров. Рассмотрены варианты с накопителями неограниченной ёмкости в обеих подсистемах, так и варианты с конечными накопителями. Для каждого из рассмотренных случаев строится двумерный случайный марковский процесс, выводится система уравнений равновесия. На основе системы уравнений равновесия получены интервальные оценки следующих характеристик: вероятность простоя всей системы, вероятность немедленного обслуживания поступающей в систему заявки, среднее число заявок в системе. Представленные интервальные верны и для случая, когда система состоит более чем из двух подсистем.

Ключевые слова: система массового обслуживания, выбор кратчайшей очереди, интервальные оценки, вероятностные и числовые характеристики.

1. Введение. Описание системы

Целью работы является исследование системы, состоящей из нескольких подсистем (накопитель и обслуживающий прибор). Поступающий в систему запрос (заявка) становится в ту подсистему, в которой находится меньше запросов (меньше длина очереди) по сравнению с другими подсистемами (Join-the-Shortest-Queue политика) [?, 1–6, 10]. Если в обеих подсистемах длины очередей совпадают, то с вероятностью p выбирается одна из подсистем, с дополнительной вероятностью $q = 1 - p$ выбирается вторая подсистема.

Реализуемый подход исследования — вместо использования асимптотических методов [4, 5], производящих функций [1–3] (не дающих точного решения для неограниченных накопителей) или метода, когда

совместное распределение вероятностей представимо как произведение частных вероятностей с некоторой нормирующей константой [7–9] (точное решение возможно при сильных упрощениях) предлагается использовать простейшие приёмы теории массового обслуживания [11] для вывода неравенств, оценивающих характеристики системы (стоит заметить, что для большинства характеристик верхняя оценка в полученных неравенствах наиболее близка к реальному значению характеристик, чем нижняя).

Предполагается, что в систему поступает пуассоновский поток интенсивности λ , времена обслуживания на приборах подчинены экспоненциальным распределениям с параметрами μ_1 и μ_2 соответственно. Ёмкости накопителей — r_1 и r_2 ($r_i \leq \infty$, $i = \overline{1, 2}$). Ряд полученных результатов совпадает с результатами работы [1, 2], актуальность исследования подчеркнута в работах [3–5], необходимый аппарат теории массового обслуживания — [11].

Функционирование системы описывается двумерным марковским процессом $\{\nu_1(t), \nu_2(t)\}$, $t \geq 0$, где $\nu_1(t)$ — случайная величина, обозначающая число заявок в системе с накопителем r_1 , а $\nu_2(t)$ — случайная величина — число заявок в системе с накопителем r_2 [1, 2, 6]. Его множество состояний имеет вид $\mathcal{X} = \{(i, j), i = \overline{0, r_1 + 1}, j = \overline{0, r_2 + 1}, r_1, r_2 \leq \infty\}$.

В предположении о существовании стационарного режима [1, 4, 6] обозначим через $p_{i,j}$ стационарную вероятность того, что в первой подсистеме находится i , $i = \overline{0, r_1 + 1}$, заявок, а во второй подсистеме — j , $j = \overline{0, r_2 + 1}$ заявок.

Через $P_{w=0}$ обозначим стационарную вероятность немедленного начала обслуживания поступившей в систему заявки:

$$P_{w=0} = \sum_{j=1}^{\infty} p_{0,j} + \sum_{i=1}^{\infty} p_{i,0} + p_{0,0}.$$

2. Интервальные оценки для системы с неограниченными накопителями

Будем рассматривать случай, когда накопители в подсистемах не ограничены. Условие существования стационарного режима приведены в работах [1, 4, 6]. Используя результаты работ [5, 7–9], можно найти вероятности простоя каждой из подсистем:

$$p_{0,\cdot} = 1 - \frac{\lambda}{\mu_1} p, \quad p_{\cdot,0} = 1 - \frac{\lambda}{\mu_2} q,$$

где $p_{0,\cdot} = \sum_{j=0}^{\infty} p_{0,j}$, $p_{\cdot,0} = \sum_{i=0}^{\infty} p_{i,0}$.

Пусть $p_k = \sum_{i=0}^k p_{i,k-i} = \sum_{j=0}^k p_{k-j,j}$ — вероятность того, что в обеих подсистемах ровно k , $k \geq 0$, заявок. Тогда для вероятности простоя системы p_0 верны следующие неравенства:

$$1 - \frac{\lambda}{\mu} \leq p_0 \leq 1 - \frac{\lambda}{2\mu}$$

для случая, когда интенсивности обслуживания совпадают ($\mu_1 = \mu_2 = \mu$), и

$$1 - \frac{\lambda}{\min(\mu_1, \mu_2)} \leq p_0 \leq 1 - \frac{\lambda}{2 \max(\mu_1, \mu_2)},$$

когда $\mu_1 \neq \mu_2$.

Вероятность немедленного обслуживания $P_{w=0}$ ограничена снизу следующей оценкой:

$$P_{w=0} \geq 1 - \frac{\lambda}{2\mu},$$

если $\mu_1 = \mu_2 = \mu$, и

$$P_{w=0} \geq 1 - \frac{\lambda}{2 \max(\mu_1, \mu_2)},$$

если $\mu_1 \neq \mu_2$.

Среднее число N заявок в системе лежит в пределах:

$$\frac{\lambda}{2\mu - \lambda} \leq N \leq \frac{\lambda}{\mu - \lambda},$$

$$\frac{\lambda}{2 \max(\mu_1, \mu_2) - \lambda} \leq N \leq \frac{\lambda}{\min(\mu_1, \mu_2) - \lambda},$$

для случаев, когда $\mu_1 = \mu_2 = \mu$ и $\mu_1 \neq \mu_2$, соответственно.

3. Интервальные оценки для систем с ограниченными накопителями

В этом разделе представлены интервальные оценки для случая, когда ёмкости накопителей ограничены — $r_1 < \infty$ и $r_2 < \infty$, причём нужно рассматривать три возможных варианта: $r_1 < r_2$, $r_1 > r_2$ (симметричен предыдущему) и $r_1 = r_2 = r$. Рассмотрим сначала вариант, когда накопители имеют равные ёмкости.

Введём вероятность p_k , $k = \overline{0, 2r+2}$, того, что в обеих подсистемах ровно k заявок:

$$p_k = \sum_{j=0}^k p_{k-j,j} = \sum_{i=0}^k p_{i,k-i}, \quad k = \overline{0, r+1},$$

$$p_k = \sum_{j=k-(r+1)}^{r+1} p_{k-j,j} = \sum_{i=k-(r+1)}^{r+1} p_{i,k-i}, \quad k = \overline{r+2, 2r+2}.$$

Оценки вероятности простоя системы p_0 и вероятности переполнения системы p_{2r+2} равны:

$$\frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{2r+2}} \leq p_0 \leq \frac{1 - \frac{\lambda}{2\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^{2r+2}},$$

$$\frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{2r+2}} \left(\frac{\lambda}{2\mu}\right)^{2r+2} \leq p_{2r+2} \leq \frac{1 - \frac{\lambda}{2\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^{2r+2}} \left(\frac{\lambda}{\mu}\right)^{2r+2},$$

если $\mu_1 = \mu_2 = \mu$, и

$$\frac{1 - \frac{\lambda}{\min(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{\min(\mu_1, \mu_2)}\right)^{2r+2}} \leq p_0 \leq \frac{1 - \frac{\lambda}{2 \max(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{2 \max(\mu_1, \mu_2)}\right)^{2r+2}},$$

$$\begin{aligned} \frac{1 - \frac{\lambda}{\min(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{\min(\mu_1, \mu_2)}\right)^{2r+2}} \left(\frac{\lambda}{2 \max(\mu_1, \mu_2)}\right)^{2r+2} &\leq p_{2r+2} \leq \\ &\leq \frac{1 - \frac{\lambda}{2 \max(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{2 \max(\mu_1, \mu_2)}\right)^{2r+2}} \left(\frac{\lambda}{\min(\mu_1, \mu_2)}\right)^{2r+2}, \end{aligned}$$

если $\mu_1 \neq \mu_2$.

Если $r_1 \neq r_2$ ($r_1 < r_2$), то стационарное распределение числа заявок в системе определяется как:

$$p_k = \sum_{i=0}^k p_{i,k-i}, \quad k = \overline{0, r_1+1},$$

$$p_k = \sum_{i=k-(r_1+1)}^{r_1+1} p_{i,k-i}, \quad k = \overline{r_1+2, 2r_1+2},$$

$$p_k = \sum_{j=r_1+1}^{k-(r_1+1)} p_{r_1+1,k-j}, \quad k = \overline{2r_1+3, r_1+r_2+2}.$$

Интервальные оценки вероятностей p_0 и $p_{r_1+r_2+1}$:

$$\frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{r_1+r_2+2}} \leq p_0 \leq \frac{1 - \frac{\lambda}{2\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^{r_1+r_2+2}},$$

$$\frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{r_1+r_2+2}} \left(\frac{\lambda}{2\mu}\right)^{r_1+r_2+2} \leq p_{r_1+r_2+2} \leq \frac{1 - \frac{\lambda}{2\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^{r_1+r_2+2}} \left(\frac{\lambda}{\mu}\right)^{r_1+r_2+2},$$

если $\mu_1 = \mu_2 = \mu$, и

$$\frac{1 - \frac{\lambda}{\min(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{\min(\mu_1, \mu_2)}\right)^{r_1+r_2+2}} \leq p_0 \leq \frac{1 - \frac{\lambda}{2 \max(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{2 \max(\mu_1, \mu_2)}\right)^{r_1+r_2+2}},$$

$$\frac{1 - \frac{\lambda}{\min(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{\min(\mu_1, \mu_2)}\right)^{r_1+r_2+2}} \left(\frac{\lambda}{2 \max(\mu_1, \mu_2)}\right)^{r_1+r_2+2} \leq p_{r_1+r_2+2} \leq \frac{1 - \frac{\lambda}{2 \max(\mu_1, \mu_2)}}{1 - \left(\frac{\lambda}{2 \max(\mu_1, \mu_2)}\right)^{r_1+r_2+2}} \left(\frac{\lambda}{\min(\mu_1, \mu_2)}\right)^{r_1+r_2+2},$$

если $\mu_1 \neq \mu_2$.

4. Заключение

В работе была рассмотрена система массового обслуживания, состоящая из двух параллельных подсистем (накопитель и обслуживающий прибор) с единым потоком заявок, поступающим в ту систему, в которой меньше заявок.

Получены интервальные оценки вероятности простоя системы, вероятности немедленного начала обслуживания поступившей в систему заявки, а также среднего числа заявок в системе

Благодарности

Работа частично поддержана грантами РФФИ № 14-07-00090, № 15-07-03007, № 15-07-03406 и № 16-07-00766.

Литература

1. *Cohen J.W., Voxxa O.J.* Boundary Value Problems in Queueing system analysis. — North-Holland, 1983.
2. *Flatto L., McLean H. P.* Two queues in parallel. // *Comm. Pure Appl. Math.* — 1977. — V. 30. — P. 255–263.
3. *Suh K., Diot Ch., Kurose J., Massoulié L., Neumann Ch., Towsley D., Varvello M.* Push-to-peer video-on-demand system: Design and evaluation. // *IEEE Journal on Selected Areas in Communications.* — 2007. — 25:9. — P. 1706–1716.
4. *Bramson M.* Stability of Join the Shortest Queue Networks. // *Ann. Appl. Probab.* — 2011. — 21:4. — P. 1568–1625.
5. *Введенская Н. Д.* Конфигурация перегруженных серверов при динамической маршрутизации. // *Пробл. передачи информ.* — 2011. — 47:3. — С. 80–95.
6. *Введенская Н. Д., Добрушин Р. Л., Карпелевич. Ф. И.* Система обслуживания с выбором наименьшей из двух очередей — асимптотический подход. // *Пробл. передачи информ.* — 1996. — 32:1. — С. 20–34.
7. *Adan I. J. B. F.* The shortest queue problem. — 2002. <http://wwwhome.math.utwente.nl/~boucherierj/onderwijs/Advanced%20Queueing%20Theory/Adan%20shortest%20queue%2019-03-2002.pdf>
8. *Adan I. J. B. F., Wessels J., Zijm W. H. M.* Analysis of the symmetric shortest queue problem. — 1990. — <https://www.researchgate.net/publication/243665959>.
9. *Adan I. J. B. F., Wessels J., Zijm W. H. M.* Analysis of the asymmetric shortest queue problem. // *Queueing Systems.* — 1991. — V. 8, No. 1, — P. 1.–58.
10. *Яно Гу Серж, Зарядов И. С.* Интервальные оценки характеристик системы с оптимальным выбором. // *Материалы Всероссийской конференции с международным участием «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологических систем».* — 2016. — РУДН. — С. 124–126.
11. *Bocharov P.P., D’Apice C., Pechinkin A.V., Salerno S.* Queueing Theory. — VSP, Utrecht, Boston, 2004.

UDC 519.218.31

Interval estimation of system performance with the optimal choice

G. S. Yapo*, T. A. Milovanova*, I. S. Zaryadov*[†]

* *Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

[†] *Institute of Informatics Problems, FRC CSC RAS
PI FRC CSC RAS, 44-2 Vavilova Str., Moscow 119333, Russia*

The queuing system consisting of two parallel subsystems (each subsystem includes a buffer and a server). The discipline of optimal choice is the following: the arriving at the system customer selects the subsystem for which the queue length in the buffer is minimal. If the queue lengths are the same sizes in both subsystems, then the incoming customer with some probability p chooses one of the subsystems and with the complementary probability $q = 1-p$ the second subsystem is selected. It is assumed that the incoming flow is a Poisson one, the time of service on the servers are exponentially distributed. The buffers in the subsystems may have unlimited capacity for both subsystems (the first considered case) and the buffers sizes may be limited (the second case). For each of these cases the two-dimensional random Markov process is constructed and the systems of equilibrium equations are derived. Based on the system of equilibrium equations we derived simple interval estimates of the following characteristics: the probability of system being idle, the probability of immediate service of incoming customers, the average number of customers in the system. Presented interval estimations are also valid for the case when the system consists of more than two subsystems.

Keywords: queuing system, join-the-shortest queue policy, interval estimates, numeric and probability characteristics.

UDC 519.872

Asymptotic Analysis of Queuing System with MMPP Arrivals and Feedback

L. Zadiranova*, A. Melikov[†], A. Moiseev*

* *Department of Probability theory and Mathematical Statistics*

Tomsk State University

Lenin ave. 36, Tomsk, 634050, Russian Federation

[†] *Laboratory of Teletraffic Theory*

Institute of Control Systems ANAS

Bakhtiyyar Vahabzadeh st. 9, Baku, Az1141, Azerbaijan

Abstract. We consider infinite-server queue with Markov modulated Poisson arrivals and feedback. Asymptotic analysis of the aggregate arrival process is made under conditions of frequent changing of the underlying chain states of the arrival process and increasing service time. It is proved that the aggregate arrival process is Poisson asymptotically. Parameter of the Poisson approximation is obtained. Applicability area of the asymptotic results is derived by means of numerical experiments. The results of the paper are similar to earlier obtained results for queueing systems with Poisson and renewal arrival processes and feedback.

Keywords: infinite-server queue, asymptotic analysis, feedback, aggregate arrival process.

1. Introduction

Queuing models have been investigated by many authors due to their different applications in production, communications, banking, computer systems and other areas. Due to present rapid development of these systems there is a need to expand the modification of queuing systems, as well as design and develop new methods of their investigation. It is known that the queuing system (QS) with feedback [1] can be used to describe socio-economic processes [2, 3], as well as afterservice processes in information and telecommunication systems [4–9].

Models of queuing systems with feedback are of two types: models with instant feedback and models with retarded feedback. In the available literature, both types of models are investigated separately.

Infinite-server queue with feedback and Poisson arrivals were considered in the articles [10, 11]. In this paper, we study the aggregate arrival process in infinite-server queue with Markov modulated Poisson arrivals and feedback.

2. Mathematical Model

Consider an infinite-server queue with arrivals as Markov modulated Poisson process (MMPP) which underlying chain $k(t)$ has a finite number of states: $1, 2, \dots, K$. The MMPP is determined by given matrix of

infinitesimals $\mathbf{Q} = \|q_{ij}\|$, $i, j = 1, 2, \dots, K$, and by matrix of conditional rates $\mathbf{\Lambda} = \text{diag}\{\lambda_k\}$, $k = 1, 2, \dots, K$ [12].

When a customer arrives at the system, it occupies any free server. Service time is distributed according to the exponential law with a parameter μ . When service is finished, the customer leaves the system with a probability $(1 - r)$ or goes into the system again for additional service with a probability r . So, the arrivals in the system is an aggregate arrival process which contains "pure" arrivals from MMPP and feedback arrivals. The problem of the study is an analysis of the aggregate arrival process.

Let us denote the following: $i(t)$ is a number of busy servers at the time instant t , $m(t)$ is a number of customers in aggregate arrival process that have come into the system during time interval $[0, t)$, $k(t)$ is a state of the underlying Markov chain of the MMPP at the time instant t . Three-dimensional process $\{k(t), i(t), m(t)\}$ is Markov. For its probability distribution $P(k, i, m, t) = P\{k(t) = k, i(t) = i, m(t) = m\}$, we can write down Kolmogorov differential equation system in the following form

$$\frac{\partial P(k, i, m, t)}{\partial t} = -\lambda_k P(k, i, m, t) - i\mu P(k, i, m, t) + \lambda_k P(k, i - 1, m, t) + \mu(i+1)(1-r)P(k, i+1, m, t) + \mu ir P(k, i, m-1, t) + \sum_{\nu=1}^K P(\nu, i, m, t)q_{\nu k} \quad (1)$$

for $k = 1, 2, \dots, K$, $i, m = 0, 1, 2, \dots$

We introduce partial characteristic functions [12].

$$H(k, u, s, t) = \sum_{i=0}^{\infty} \sum_{m=0}^{\infty} e^{ju_i} e^{j s m} P(k, i, m, t),$$

where $j = \sqrt{-1}$. Then we can write the system (1) in form of differential matrix equation

$$\frac{\partial \mathbf{H}(u, s, t)}{\partial t} + j\mu [r e^{js} - 1 + (1-r)e^{-ju}] \frac{\partial \mathbf{H}(u, s, t)}{\partial u} = \mathbf{H}(u, s, t) \left[\left(e^{j(u+s)} - 1 \right) \mathbf{\Lambda} + \mathbf{Q} \right], \quad (2)$$

where $\mathbf{H}(u, s, t) = [H(1, u, s, t), \dots, H(K, u, s, t)]$.

3. Asymptotic Analysis

Consider the system MMPP/M/ ∞ described above. Let us fix a matrix of infinitesimals as $\mathbf{Q}^{(1)}$. Let S be a positive value. Denote $\mathbf{Q} = S \cdot$

$\mathbf{Q}^{(1)}$. It is clear that the stationary probability distributions of state of the underlying Markov chain $k(t)$ coincide for the cases both with infinitesimal matrices $\mathbf{Q}^{(1)}$ and \mathbf{Q} . In other words, this distribution does not depend on parameter S , but the intensity of the transitions of Markov chain $k(t)$ from one state to another increases while value of the parameter S increases. This corresponds us to the condition of frequent changing of the underlying chain states of the arrival process.

We denote

$$\mu = \varepsilon, u = \varepsilon y, \frac{1}{S} = \varepsilon, \mathbf{H}(u, s, t) = \mathbf{F}(y, s, t, \varepsilon).$$

Let us rewrite the equation (2), taking into account these notations:

$$\frac{\partial \mathbf{F}(y, s, t, \varepsilon)}{\partial t} = j\varepsilon [1 - re^{js} - (1 - r)e^{-j\varepsilon y}] \frac{\partial \mathbf{F}(y, s, t, \varepsilon)}{\varepsilon \cdot \partial y} + \mathbf{F}(y, s, t, \varepsilon) \left[\left(e^{j(\varepsilon y + s)} - 1 \right) \mathbf{\Lambda} + S\mathbf{Q}^{(1)} \right]. \tag{3}$$

Let us obtain the steady-state asymptotic characteristic function of the number of busy servers under conditions of frequent changing of the underlying chain states of the arrival process and increasing service time. To do this, we set $s = 0$ and make a transition to the steady state ($t \rightarrow \infty$), then we derive the following equation

$$j(1 - r) (e^{-j\varepsilon y} - 1) \frac{\partial \mathbf{F}(y, \varepsilon)}{\partial y} = \mathbf{F}(y, \varepsilon) \left[\left(e^{j(\varepsilon y)} - 1 \right) \mathbf{\Lambda} + S\mathbf{Q}^{(1)} \right].$$

Let us investigate the number of customers in the aggregate arrival process under conditions of frequent changing of the underlying chain states of the arrival process and increasing service time. Denote by \mathbf{e} a column vector which consists of 1's.

Theorem 1 *Let \mathbf{R} be a vector of stationary probability distribution of states of Markov chain $k(t)$ which is determined by the system of equations*

$$\begin{cases} \mathbf{R}\mathbf{e} = 1, \\ \mathbf{R}\mathbf{Q} = \mathbf{0}. \end{cases}$$

Then the asymptotic characteristic function of the number of customers in the aggregate arrival process $h(s, t) = M \{e^{j\mathbf{sm}(t)}\}$ has the form

$$h(s, t) = \exp \left\{ \frac{\kappa (e^{js} - 1)}{1 - r} \right\}$$

under the conditions of frequent changing of the underlying chain states of the arrival process and increasing service time. Here $\kappa = \mathbf{R}\mathbf{\Lambda}\mathbf{e}$.

Proof. Dividing right and left side of the equation (3) by S and making a transition $S \rightarrow \infty$, we obtain the system of equations

$$\mathbf{0} = \mathbf{F}(u, s, t)\mathbf{Q}^{(1)}.$$

Its solution has the form

$$\mathbf{F}(u, s, t) = \mathbf{R}\Phi(u, s, t), \tag{4}$$

where $\Phi(u, s, t)$ is a some scalar function.

To determine the form of this function, we multiply both sides of the equation (3) by the vector \mathbf{e} :

$$\frac{\partial \mathbf{F}(y, s, t, \varepsilon)}{\partial t} \mathbf{e} = j [1 - re^{js} - (1 - r)e^{-j\varepsilon y}] \frac{\partial \mathbf{F}(y, s, t, \varepsilon)}{\partial y} \mathbf{e} + \mathbf{F}(y, s, t, \varepsilon) \left[\left(e^{j(\varepsilon y + s)} - 1 \right) \mathbf{\Lambda} + S\mathbf{Q}^{(1)} \right] \mathbf{e}.$$

Performing an asymptotic transition $\varepsilon \rightarrow 0$ and taking into account the form (4), we derive the following differential equation

$$\frac{\partial \Phi(y, s, t)}{\partial t} = jr (1 - e^{js}) \frac{\partial \Phi(y, s, t)}{\partial y} + \Phi(y, s, t)\kappa (e^{js} - 1).$$

Solving this equation under the initial condition [13] $\Phi(y, s, 0) = \exp \left\{ \frac{j\kappa y}{1 - r} \right\}$, we obtain

$$\Phi(y, s, t) = \exp \left\{ \frac{\kappa t (e^{js} - 1)}{1 - r} + \frac{j\kappa y}{1 - r} \right\}.$$

Therefore, the solution of equation (3) has the form

$$\mathbf{F}(y, s, t) = \mathbf{R} \exp \left\{ \frac{\kappa t (e^{js} - 1)}{1 - r} + \frac{j\kappa y}{1 - r} \right\}.$$

Assuming here $y = 0$, we derive the following asymptotic approximation of the characteristic function of the number of customers in the aggregate arrival process under conditions of frequent changing of the underlying chain states of the arrival process and increasing service time:

$$h(s, t) = M \left\{ e^{jsm(t)} \right\} = \mathbf{H}(0, s, t)\mathbf{e} \approx \mathbf{F}(0, s, t)\mathbf{e} = \exp \left\{ \frac{\kappa t (e^{js} - 1)}{1 - r} \right\}.$$

The theorem is proved.

This theorem shows that the aggregate arrival process in the system has the Poisson distribution with parameter $\frac{\kappa t}{1 - r}$.

4. Conclusions

In the paper, an asymptotic characteristic function of the number of customers in the aggregate arrival process is obtained for infinite-server queue with MMPP arrivals and a feedback under conditions of frequent changing of the underlying chain states of the arrival process and increasing service time. Obtained numerical results show that the range of applicability of the asymptotic method is increased in comparison to the same model without frequently changing of the underlying chain states of the MMPP arrivals [14].

Some more complicate asymptotic analysis for queueing networks with MAP arrivals, but only about number of busy servers, can be found in [15].

References

1. Melikov A. Z., Ponomarenko L. A., Kuliyeva K. H. N. Calculation of the characteristics of multichannel queuing system with pure losses and feedback // *J. Autom. Inf. Sci.* — Vol. 47, no. 5. — P. 19–29.
2. Morozova A. S., Moiseeva S. P., Nazarov A. A. Investigation of economic-mathematical model of the effect of price discounts for regular customers profit commercial organization // *Tomsk State University Journal (in Russian)*. — Vol. 293. — P. 49–52.
3. Zakhornolnaya I. A., Moiseeva S. P. Mathematical model of the process of changes in income from the sale of complementary products // *Financial and Actuarial Mathematics and eventokonvergensii technologies (FAMET – 2011): Proceedings of the X International Conference: 23–24 april, 2011. (in Russian)*. — P. 157–170.
4. Wan-Chun C. A computer processing queueing system with feedback // *Information and Control*. — Vol. 16. — P. 473–486.
5. Foley R. D., Disney R. L. Queues with delayed feedback // *Adv. Appl. Probab.* — Vol. 15, no. 1. — P. 162–182.
6. Pekoz E. A., Joglekar N. J. Poisson traffic flow in a general feedback // *Appl. Probab.* — Vol. 39, no. 3. — P. 630–636.
7. Zaryadov I., Korolkov A., Milovanov T., Sherbanskaya A. Mathematical model of calculating and analysis of characteristics of systems with generalized and repeating service // *T-Comm – Telecommunications and Transport (in Russian)*. — Vol. 8, no. 6. — P. 16–20.
8. Zaryadov I., Sherbanskaya A. Time characteristics of queueing system with renovation and reserve // *Bulletin of PFUR. Series Mathematics. Information Sciences. Physics*. — Vol. 2. — P. 61–65.
9. D'Avignon G. R., Disney R. L. Queues with instantaneous feedback // *Manag. Sci.* — Vol. 24.
10. Morozova A. S., Moiseeva S. P., Nazarov A. A. Research of qs with repeating requests and unlimited number of servers by means of a method of limit decomposing // *Computing technologies (in Russian)*. — Vol. 5, no. 13. — P. 88–92.

11. Morozova A. S., Moiseeva S. P., Nazarov A. A. Research of qs with repeating requests and unlimited number of servers by means of a method of limit decomposing // Tomsk State University Journal (in Russian). — Vol. 293. — P. 49–52.
12. Nazarov A. A., Moiseeva S. P. Methods of asymptotic analysis in theory of mass queuing. — Tomsk: Publ. NTL (in Russian), 2006.
13. Zhidkova L. A., Moiseeva S. P. Investigation of occupied servers in the queueing system $m\text{mpp}/m/\infty$ with feedback // Tomsk State University Journal (in Russian). — Vol. 1, no. 26. — P. 53–62.
14. Zadiranova L. A., Moiseeva S. P. Investigation of the aggregate arrival process of qs with feedback by the method of asymptotic analysis // Proceedings of the Tomsk State University (in Russian). — Vol. 297. — P. 99–105.
15. Moiseev A., Nazarov A. Queueing network $\text{map}-(\text{gi}-\infty)\text{k}$ with high-rate arrivals // European Journal of Operational Research DOI: 10.1016/j.ejor.2016.04.011. — Vol. 254. — P. 161–168.

УДК 519.218.31

Система массового обслуживания с обобщённым обновлением и повторным обслуживанием — вероятностно-временные характеристики

И. С. Зарядов*[†], И. А. Мацкевич*, А. А. Щербанская*

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

[†] *Институт проблем информатики ФИЦ ИУ РАН,
ул. Вавилова 44 кор.2, ИПИ ФИЦ ИУ РАН, Москва, Россия, 119333*

Аннотация. Исследуется система массового обслуживания, состоящая из накопителя неограниченной ёмкости и одного прибора с экспоненциально распределённым временем обслуживания заявок на нем. В систему поступает рекуррентный поток заявок с конечным средним значением времени между последовательными моментами поступления. Вводится дисциплина обобщённого обновления с повторным обслуживанием (сразу после окончания своего обслуживания на приборе заявка либо с некоторой вероятностью покидает систему, либо с иной вероятностью сбрасывает из накопителя произвольное число заявок и переходит в накопитель, становясь в конец очереди). Получены в аналитическом виде выражения для расчёта основных вероятностно-временных характеристик рассматриваемой системы.

Ключевые слова: обобщённое обновление, повторное обслуживание, вероятностные и временные характеристики.

1. Введение

Интерес к рассматриваемому классу систем массового обслуживания связан с развитием компьютерных и телекоммуникационных систем, и впервые были рассмотрены А.Я. Крейниным [1].

В работе исследуется однолинейная система массового обслуживания с накопителем бесконечной ёмкости. Входящий в систему поток является рекуррентным, причём время между соседними поступлениями заявок имеет произвольную функцию распределения $A(x)$.

Обобщённое обновление [3, 4] с повторным обслуживанием [2, 6] определяется следующим образом: в момент окончания своего обслуживания с некоторой вероятностью $q(i)$, $i \geq 0$, заявка на приборе сбрасывает из накопителя i других заявок и становится снова в очередь, либо с вероятностью p просто покидает систему. Кроме того, через q обозначим вероятность того, что обслужившаяся заявка полностью опустошит накопитель независимо от того, сколько других заявок было в накопителе, и останется в системе [6]. Системы с обобщённым обновлением схожи с системами с отрицательными заявками [7, 8]

Применение результатов для систем с обновлением показано в работе [5].

2. Вероятностные характеристики

При исследовании данной системы будем опираться на уже известные результаты для систем с обновлением с повторным обслуживанием [2, 6], так и для систем с обобщённым обновлением без повторного обслуживания [3, 4].

Исследование данной системы проведём с помощью вложенной цепи Маркова (см., например, [9]), образованной числами $\nu(\tau_n - 0)$ заявок в системе в моменты времени $\tau_n - 0$, где τ_n — момент поступления n -й заявки. Множество состояний вложенной цепи Маркова имеет вид $\mathcal{X} = \{1, \dots, \infty\}$.

Пусть $p_i, i \geq 0$, стационарная вероятность того, что поступающая в систему заявка застанет в ней ровно i других заявок. Стационарное распределение вероятностей для систем подобного вида имеет геометрический вид [3, 4, 6]: Условия существования стационарного режима в системах подобного вида приведены в [1, 3].

Для дисциплины обобщённого обновления с повторным обновлением помимо стационарного распределения числа заявок в системе и вероятности обслуживания (потери) поступающей заявки (основные вероятностные характеристики для систем массового обслуживания с обобщённым обновлением [3, 4]) можно выписать ряд дополнительных вероятностных характеристик.

Вероятность π_1 того, что все поступающие в систему заявки будут обслужены (ни одна из них не была сброшена из накопителя) и все они покинут систему, определяется следующим образом:

$$\pi_1 = \frac{pp_1}{1 - gp}.$$

Вероятность быстрого переполнения системы π_2 , то есть вероятность того, что ни одна из закончивших обслуживание заявок не покинет систему, и при этом из накопителя не будет сброшено ни одной заявки (заявки не покидают систему) имеет вид:

$$\pi_2 = \frac{p_1q(0)}{1 - gp}.$$

Вероятность π_3 — вероятность того, что ни одна заявка в системе не будет потеряна (ни одна заявка не сброшена из накопителя) представима в виде:

$$\pi_3 = \frac{p_1(p + q(0))}{1 - g(p + q(0))}.$$

Для расчёта вероятности того, что произвольная поступающая в систему заявка не будет потеряна (сброшена из накопителя) и, соответственно, будет обслужена на приборе, вводится вспомогательная производящая функция $\pi_{1,1}(u) = \sum_{j=1}^{\infty} u^j q(j)$.

Вероятность $p^{(serv)}$ того, что заявка, поступающая в систему, перейдёт на обслуживание, вычисляется по формуле:

$$p^{(serv)} = \frac{p_1(p + \pi_{1,1}(g))}{(1-g)(1-g(p + \pi_{1,1}(g)))} + p_0 + \pi_1.$$

Очевидно, что вероятность потери поступившей заявки $p^{(loss)}$ вычисляется как дополнительная к $p^{(serv)}$ вероятность: $p^{(loss)} = 1 - p^{(serv)}$.

3. Временные характеристики системы

В качестве основной временной характеристики системы будем рассматривать стационарное распределение времени ожидания начала обслуживания принятой (и, в дальнейшем, не сброшенной) заявки.

Сделаем два предположения: заявки обслуживаются в порядке поступления, кроме того, заявки, закончившие обслуживание и оставшиеся в системе, становятся в конец очереди в накопителе.

Обозначим через $W^{(serv)}(x)$ стационарную функцию распределения времени ожидания начала обслуживания принятой к обслуживанию заявки:

$$W^{(serv)}(x) = \frac{1}{p^{(serv)}} \left(p_0 + \sum_{i=1}^{\infty} p_i W_i^{(serv)}(x) \right),$$

где $W_i^{(serv)}(x)$ — вероятность того, что время ожидания заявки, поступившей в систему и заставшей в ней i других заявок ($i \geq 1$), будет меньше x при условии, что эта заявка не была сброшена и перешла на обслуживание.

Введём преобразование Лапласа-Стилтьеса (ПЛС) $\omega^{(serv)}(s)$ функции $W^{(serv)}(x)$. Тогда в терминах ПЛС стационарная функция распределения времени ожидания начала обслуживания принятой к обслуживанию заявки имеет вид:

$$\omega^{(serv)}(s) = \frac{1}{p^{(serv)}} \left(p_0 + \frac{p_1 \mu}{1-g} \left(\frac{p}{s + \mu(1-gp)} + \frac{p + \pi_{1,1}(g)}{s + \mu(1-g(p + \pi_{1,1}(g)))} \right) \right).$$

Из ПЛС $\omega^{(serv)}(s)$ получаем среднее время ожидания начала обслуживания

$$w^{(serv)} = \frac{p_1 \mu}{p^{(serv)}(1-g)} \left(\frac{p}{(\mu(1-gp))^2} + \frac{p + \pi_{1,1}(g)}{(\mu(1-g(p + \pi_{1,1}(g))))^2} \right)$$

4. Заключение

В работе была исследована система массового обслуживания с обобщённым обновлением и повторным обслуживанием. Для неё получены основные вероятностно-временные характеристики — стационарное распределение числа заявок в системе, вероятность обслуживания поступающей заявки, вероятность быстрого переполнения системы, распределение времени пребывания в накопителе принятой к обслуживанию заявки, а также среднее время ожидания начала обслуживания поступившей заявки.

Благодарности

Работа частично поддержана грантами РФФИ № 14-07-00090, № 15-07-03007, № 15-07-03406 и № 16-07-00766.

Литература

1. *Kreinin A.* Queueing Systems with Renovation. // Journal of Applied Math. Stochast. Analysis. — 1997. — Vol. 10, No. 4. — P. 431–443.
2. *Бочаров П. П., Зарядов И. С.* Стационарное распределение вероятностей в системах массового обслуживания с обновлением. // Вестник РУДН, серия «Математика. Информатика. Физика». — 2007. — № 1-2. — С. 14–23.
3. *Zaryadov I. S., Pechinkin A. V.* Stationary Time Characteristics of the GI/M/n/ System with Some Variants of the Generalized Renovation Discipline. // Automation and Remote Control. — 2009. — No. 12. — P. 2085–2097.
4. *Zaryadov I. S.* The GI/M/n/ Queuing System with Generalized Renovation. // Automation and Remote Control. — 2010. — No. 4. — P. 663–671.
5. *Зарядов И. С., Королькова А. В.* Применение модели с обобщенным обновлением к анализу характеристик систем активного управления очередями типа Random Early Detection (RED). // Т-Comm. Коммуникации и транспорт. — 2011. — № 7. — С. 84–88.
6. *Зарядов И. С., Королькова А. В., Милованова Т. А., Щербанская А. А.* Математическая модель расчёта и анализа характеристик систем с обобщенным обновлением и повторным обслуживанием // Т-Comm: Телекоммуникации и транспорт. — 2014. — № 6. — С. 26–20.
7. *Gelenbe E., Glynn P., Sigman K.* Queues with negative arrivals // J. Appl. Prob. — 1991. — V. 28. — P. 245–250.
8. *Бочаров П. П., Д'Анжич Ч., Мандзо Р., Печинкин А. В.* Анализ многолинейной марковской системы массового обслуживания с неограниченным накопителем и отрицательными заявками // Автоматика и телемеханика. — 2007. — № 1. — С. 93–104.

9. Bocharov P.P., D'Apice C., Pechinkin A.V., Salerno S. Queueing Theory. — VSP, Utrecht, Boston, 2004.

UDC 519.218.31

The queueing system with general renovation and repeated service — time-probability characteristics

I. S. Zaryadov^{*†}, I. A. Matskevich^{*}, A. A. Scherbanskaya^{*}

** Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

*† Institute of Informatics Problems
Federal Research Center "Computer Science and Control" of Russian Academy
of Sciences
PI FRC CSC RAS, 44-2 Vavilova Str., Moscow 119333, Russia*

The queueing system, consisting of an unlimited buffer and a single server with exponentially distributed service time, with general input flow is under investigation. In addition, discipline of general renovation with repeated service (this discipline means that the served customer may leave the system with some probability or may drop arbitrary number of customers from the buffer with some probability and then return to the buffer for repeated service) is introduced. The steady-state probability distribution of number of customers in the system is derived. The probability of arriving customer to be lost due to general renovation discipline and the probability of arriving customer to be served are obtained in analytical form. In addition, the probability of quick overflow of the system (none of the customers leaves the system), the probability that none of the customers will be dropped from the system are also presented. The steady-state waiting time distribution is also obtained and presented in the work.

Keywords: general renovation, repeated service, time-probability characteristics.

УДК 004.4

Кинетический подход в моделях генерации нестационарных процессов случайного блуждания и расчета функционалов на них

С. Л. Федоров

*Российский университет дружбы народов, Москва
117198, Москва, ул. Милухо-Маклая, д. 6*

Аннотация. Разработан метод генерации ансамбля траекторий нестационарного временного ряда, выборочные плотности распределения которого эволюционируют в соответствии с уравнением Фоккера-Планка. Параметры сноса и диффузии в этом уравнении оцениваются по фактическим выборкам фрагментов временного ряда, причем выборочные распределения остаются в классе распределений, характерных для данного ряда. Подчеркнем, что класс распределений выделяется непараметрическими методами.

Ключевые слова: нестационарный временной ряд, моделирование ансамбля траекторий, тестирование функционала.

1. Введение

Проблема моделирования нестационарных временных рядов возникает во многих областях человеческой деятельности. Существует множество примеров рядов данных, требующих моделирования с учетом нестационарных свойств выборочных распределений наблюдаемых величин. Таковыми являются биржевые ряды цен сделок на финансовые инструменты, кардиограммы и энцефалограммы в медицине, сейсмограммы, температурные кривые и показатели счетчиков радиоактивности, последовательности символов в текстах и цепочках геномов. Традиционный подход к анализу нестационарных временных рядов состоит в том, что рассматриваются только такие ряды, которые с помощью различных линейных методов можно свести к стационарным. Однако на практике такой подход не всегда годится, потому что некоторые наблюдаемые эффекты кроются именно в нестационарном поведении временного ряда. В данной работе будет продемонстрирован алгоритм генерации ансамбля траекторий с учетом нестационарных свойств выборочных распределений наблюдаемых величин.

Исследование выполнено в рамках соглашения № 02.а03.21.0008 от 24.04.2016 г. между Министерством образования и науки Российской Федерации и Российским университетом дружбы народов.

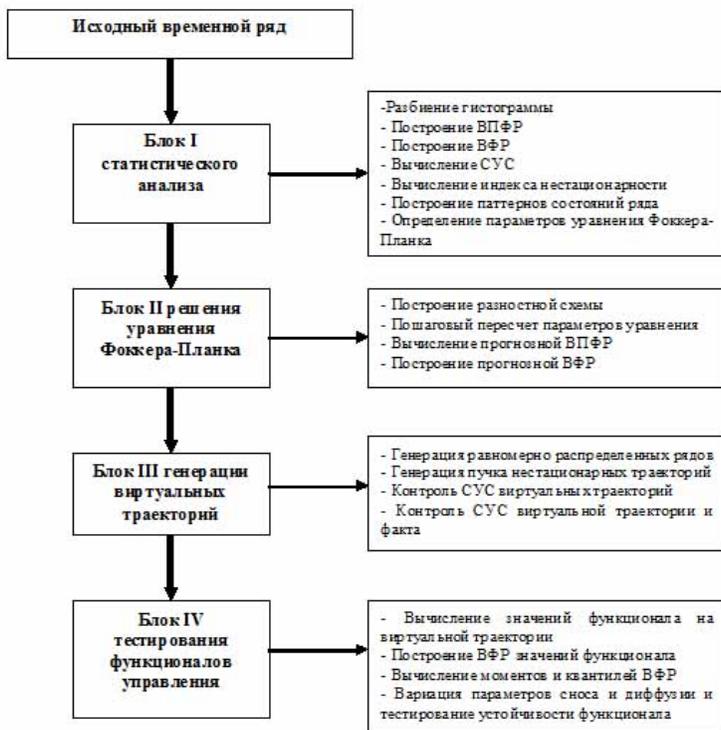


Рис. 1. Структура программного алгоритма

2. Метод генерации нестационарных случайных траекторий

Итак, создан программный комплекс, позволяющий анализировать и прогнозировать нестационарные временные ряды. Он состоит из четырех последовательных блоков 1. На вход подается исходный временной ряд. В результате работы программы пользователь получает ансамбль траекторий с нестационарными свойствами выборочных распределений характерных для исходного временного ряда. Используя полученный ансамбль, пользователь может рассчитать значение некоторого заданного функционала в широком диапазоне выборочных статистик.

3. Заключение

В работе описан метод генерации траекторий нестационарных временных рядов, имитирующий временной ряд, возникающий на практике в процессе наблюдения за случайной величиной, распределение которой не является стационарным. Моделирование ряда основано на решении эмпирического уравнения Фоккера-Планка, причем выборочные распределения остаются в классе распределений, характерных для данного ряда. Подчеркнем, что класс распределений выделяется непараметрическими методами. Описанный метод позволяет тестировать функционалы, заданные на случайном временном ряде в широком диапазоне изменения выборочных статистик ряда.

Благодарности

Исследование выполнено в рамках соглашения № № 02.а03.21.0008 от 24.04.2016 г. между Министерством образования и науки Российской Федерации и Российским университетом дружбы народов.

Литература

1. *Ванник В. Н., Червоненкис А. Я.* Теория распознавания образов. — М.: Наука, 1974, 416 с.
2. *Кобзарь А. И.* Прикладная статистика. — М.: Физматлит, 2006, 816 с.
3. *Орлов Ю. Н., Шагов Д. О.* Индикативные статистики для нестационарных временных рядов. / Препринты ИПМ им. М.В. Келдыша, №53, 2011, 20 с.
4. *Орлов Ю. Н., Федоров С. Л., Давидько В. А.* К вопросу классификации нестационарных временных рядов: состав индекса РТС. // Препринты ИПМ им. М.В. Келдыша, №54, 2014, 18 с.
5. *Орлов Ю. Н., Федоров С. Л.* Моделирование и статистический анализ функционалов, заданных на выборках из нестационарного временного ряда. // Препринты ИПМ им. М.В. Келдыша, №43, 2014, 26 с.
6. *Босов А. Д., Кальметьев Р. Ш., Орлов Ю. Н.* Эмпирическое уравнение Фоккера-Планка для прогнозирования нестационарных временных рядов. // Препринты ИПМ им. М.В. Келдыша, №3, 2013, 30 с.
7. *Королук В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф.* Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985, 640 с.
8. *Орлов Ю. Н.* Кинетические методы исследования нестационарных временных рядов. — М.: МФТИ, 2014, 276 с.
9. *Орлов Ю. Н., Осминин К. П.* Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков. — М.: Эдиториал УРСС, 2011, 384 с.

UDC 004.4

Kinetic approach in models of forecasting non-stationary time-series and functional calculations on them

S. L. Fedorov

*RUDN University
117198, Moscow, Miklukho-Maklaya str. 6, Russia*

The method of non-stationary time-series trajectory generation is proposed in accordance with Fokker-Plank equation for the empirical distribution function density. Parameters of trend and diffusion are estimated on the samples of time-series. The numerical algorithm for pattern recognition functional testing in the non-stationary probability conditions is constructed.

Keywords: non-stationary time-series, trajectories modeling, pattern recognition functional testing .

UDC 621.39

The Estimation of Probability Characteristics of Cloud Computing Systems with Splitting of Requests

A. V. Gorbunova*[†], I. S. Zaryadov*[†],
S. I. Matushenko*, E. S. Sopin*[†]

* *Department of Applied Probability and Informatics
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

[†] *Institute of Informatics Problems
Federal Research Center "Computer Science and Control"
Russian Academy of Sciences
Vavilova str. 44-2, Moscow, 119333, Russia*

Abstract. Growing popularity of cloud services is explained by many advantages of them. The accessibility, flexibility, scalability, ease of management, the relatively low cost of implementation can be listed among the main advantages. The demand for cloud services with the ability to change one cloud service provider to another one without any significant cost for a user result in a high competition between cloud providers. Due to this reason, it became important to find the optimal performance measures of cloud systems. These measures, on the one hand, must meet all the requirements of Service Level Agreement (SLA), on the other hand, do not lead to excessive costs for provider. The paper presents the evaluation of the main service quality characteristics of cloud systems, including formulas for variance of residence time in the synchronization buffer. For the analysis of a cloud system, fork-join queues with corresponding methods of its approximation were used.

Keywords: cloud computing, fork-join queuing system, response time, synchronization buffer, synchronization time.

1. Introduction

The interest in cloud systems, in addition to the main advantages of their use, is defined by the need of commercial and research organizations in a powerful computing infrastructure to perform resource-intensive tasks at relatively low cost [1]. The one of the most important indicators of quality of service for the cloud system is the response time. In the context of the decision of cluster problems, that are characterized by a high level of parallelism, the request is considered to be executed immediately after the processing of the last of its constituent sub-queries, so the response time is the maximum of staying time of sub-queries in the system. In addition, the one of the most effective ways to reduce the response time is to send the same task to several computing machines, and then - to wait for the earliest response. In this case, the response time is the minimum of the staying time of all sub-queries (subtasks) in the system [2]. Along with such an important quality of service metrics as the system response time, in recent years the time spent by sub-queries in the synchronization buffer (i.e. the

time between the end of the first and last sub-queries of the same task), became an essential parameter for the parallel computing [3]. This trend is explained by the fact that the longer the query (task) assembly time, the greater should be the size of the buffer. In other words, until the last sibling sub-query of the query is processed all its components will be located in the buffer with sub-queries of next queries received by the system. As the number of computers or types of sub-queries increase, also the buffer size should increase. In this context, it important to select the optimal values of system parameters functioning to minimize the synchronization time without sacrificing the quality of service.

2. The model of a cloud center

The functioning of a cloud computing center can be described in terms of a fork-join queueing system with K $M/M/1$ -type subsystems with homogeneous servers [4] (Fig. 1). Let us denote ξ_k - the random value of residence time of the k -th type sub-query (the k -th part of query).

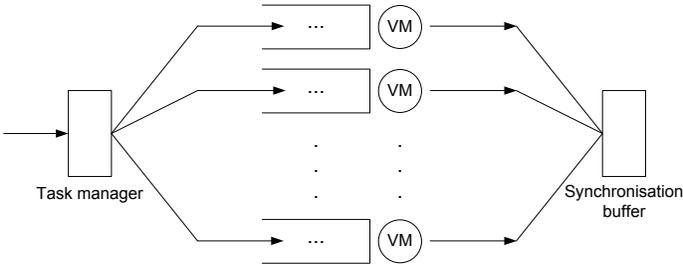


Figure 1. The queueing system model of the cloud center

We should note that the residence times of sub-queries are dependent random variables due to the joint moment of arrival. Because the exact solution for fork-join system with $K > 2$ is unknown and the possibility of obtaining this solution is under question [5, 8], we will approximate our system by K parallel independent functioning queueing systems of $M_\lambda/M_\mu/1$ type. We can use such approximation because the marginal probability distributions of k type sub-queries, $k = \overline{1, K}$, for fork-join system are similar with probability distributions for the queueing system $M/M/1$ [6].

3. Probability characteristics of the system

In this section, we consider the probability characteristics of random variables that determine performance measures of a cloud center.

The table 1 shows the random variables that characterize the main performance metrics. In order to evaluate and use these metrics when choosing a cloud service provider, we need to obtain expressions for the basic probability characteristics of random variables, i.e. the expressions for their mean and variance.

The main results for mentioned above characteristics, obtained in [5], [7] and [8] for $K \geq 2$:

$$EW_{K,\max} \approx \left[\frac{H_K}{H_2} + \frac{4}{11} \left(1 - \frac{H_K}{H_2} \right) \rho \right] \left(\frac{12 - \rho}{8} \right) \frac{1}{\mu(1 - \rho)}, \quad (1)$$

$$EW_{K,\max} \approx \left[H_K + \left(\left(\sum_{k=1}^K \binom{K}{k} (-1)^{k-1} \sum_{i=1}^k \binom{k}{i} \frac{(i-1)!}{k^{i+1}} \right) - H_K \right) \rho \right] \frac{1}{\mu - \lambda}, \quad (2)$$

$$EW_{K,\max} \approx \frac{1}{\mu} \left(H_K + \frac{\rho}{2(1 - \rho)} \left(\sum_{k=1}^K \frac{1}{k - \rho} + (1 - 2\rho) \sum_{k=1}^K \frac{1}{k(k - \rho)} \right) \right), \quad (3)$$

$$EW_{K,\max} \approx \frac{1}{\mu - \lambda} H_K, \quad (4)$$

Table 1
The random variables that determine performance measures of a cloud system

Performance metrics	Definition	Random variable	Notation in the order statistics theory
The response time	The staying time of the last served sub-query.	$W_{K,\max} = \max(\xi_1, \dots, \xi_K)$	$\xi_{(K)}$
	The staying time of the first served sub-query	$W_{K,\min} = \min(\xi_1, \dots, \xi_K)$	$\xi_{(1)}$
The synchronization time	The time between the end of the service of the first and the last sub-queries	$W_K = W_{K,\max} - W_{K,\min}$	$W_K = \xi_{(K)} - \xi_{(1)}$ (range)

where $H_K = \sum_{k=1}^K \frac{1}{k}$ - is a partial sum of the harmonic series, and $\rho = \frac{\lambda}{\mu} < 1$. The upper border of an approximation for the maximum of independent identically distributed random variables, as shown in [9], is $\xi(K) \leq \frac{1}{(\mu-\lambda)}(1 + \frac{K-1}{\sqrt{2K-1}})$. The expression for the variance of the response time is given in [4, 6]:

$$DW_{K,\max} \approx \frac{1}{(\mu - \lambda)^2} \left(2 \sum_{k=1}^K \binom{K}{k} (-1)^{k-1} \frac{1}{k^2} - H_K^2 \right). \quad (5)$$

The random variable $W_{K,\min}$ has the following characteristics:

$$EW_{K,\min} \approx \frac{1}{(\mu - \lambda)K}, \quad DW_{K,\min} \approx \frac{1}{(\mu - \lambda)K^2}.$$

For the synchronization time the following formulas are valid:

$$EW_K \approx \frac{1}{(\mu - \lambda)} H_{K-1}, \quad DW_K \approx \frac{1}{(\mu - \lambda)} \sum_{k=1}^{K-1} \frac{1}{k^2}.$$

Keep in mind that $\lim_{K \rightarrow \infty} DW_K = \frac{\pi^2}{6(\mu-\lambda)^2}$.

4. The numerical example

In order to illustrate the quality of approximation of formulas (1), (2), (3), (4) the diagrams of mean response time dependence on the number K of sub-queries are presented in Fig. 2. The calculations were performed for the arrival rate $\lambda = 2 \text{ c}^{-1}$ and service rate $\mu = 2.2 \text{ c}^{-1}$. Since for $K = 2$ the exact expression of mean response time (5) is obtained, the number of sub-queries $K > 2$ was considered.

As can be seen from the graphs, the formulas (1), (2), (3), (4) are presented in descending order of approximation accuracy, and the upper bound for the mean value of the maximum of order statistics is the upper bound for the mean response time.

The mean value and standard deviation of the synchronization time are presented in Fig. 3. One can see that the estimation of the mean synchronization time has a low accuracy contrary to the estimation of the standard deviation (the relative error is 7% on average).

5. Conclusions

In this paper, the results for the evaluation of main cloud computing systems performance measures are presented. Our further research directions include improvement of given approximations.

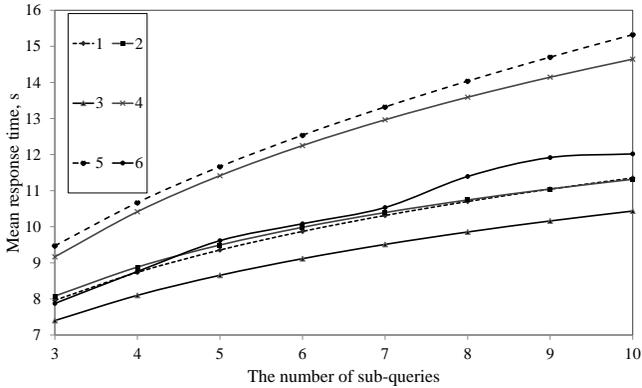


Figure 2. Mean response time: 1 – formula (1), 2 – formula (2), 3 – (3), 4 – (4), 5 – the upper bound for the maximum of order statistics, 6 – imitation model

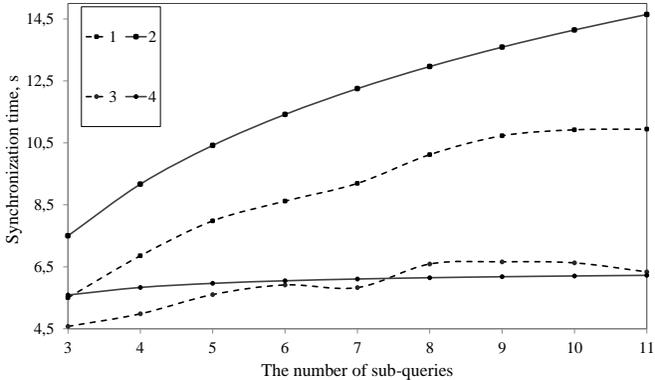


Figure 3. The mean synchronization time: 1 – the simulation model, 2 – the analytical model; the standard deviation of the synchronization time: 3 – the simulation model, 4 – the analytical model

Acknowledgments

This work is partially supported by the Russian Foundation for Basic Research (grants No 15-07-03051,15-07-03608).

References

1. *Khazae H., Mistic J., Mistic V. B.* A fine-grained performance model of cloud computing centers // IEEE Transactions on Parallel and Distributed Systems — No. 24.
2. *Joshi G., Soljanin E., Wornell G.* Efficient redundancy techniques for latency reduction in cloud systems — Unpublished.
3. *Tsimashenka I., Knottenbelt W. J.* Reduction of subtask dispersion in fork-join systems. // The 10th European Performance Evaluation Workshop (EPEW'13).
4. *Thomasian A.* Analysis of fork/join and related queueing systems. — Vol. 47, no. 17. — P. 1–71.
5. *Nelson R., Tantawi A. N.* Approximate analysis of fork/join synchronization in parallel queues. — Vol. 37. — P. 739–743.
6. *Gorbunova A. V., Zaryadov I. S., Matyushenko S. I., Samouylov K. E., Shorgin S. Ya.* The approximation of response time of a cloud computing system. — Vol. 9, — P. 32–38.
7. *Varma S., Makowski A. M.* Interpolation approximations for symmetric fork-join queues. Vol. 20. — P. 245–265.
8. *Varki E., Merchant A., Chen H.* The M/M/1 fork-join queue with variable subtasks. — Unpublished.
9. *David H. A.* Order Statistics. — New York: Wiley, 1981.

УДК 519.217

Анализ чувствительности характеристик надёжности модели дублированной системы передачи данных к виду распределений времени безотказной работы и ремонта её элементов

Г. Ж. К. Уанкпо*, Д. В. Козырев*†

* *Кафедра прикладной информатики и теории вероятностей,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

† *Институт проблем управления им. В.А. Трапезникова РАН,
Профсоюзная ул., 65, Москва, Россия*

Аннотация. Рассматривается система $\langle M_2/GI/1 \rangle$ холодного резервирования с экспоненциальной функцией распределения (ФР) времени безотказной работы (в.б.р.) и произвольной ФР времени ремонта её элементов. Получены явные аналитические выражения для стационарного распределения вероятностей состояний системы и для стационарного в.б.р. системы. Проведено имитационное моделирование системы на основе дискретно-событийного подхода. Численные и графические результаты имитационного моделирования хорошо аппроксимируют аналитическую модель системы, а значит она может быть использована в тех случаях, когда не удается получить выражения для стационарных вероятностей состояний системы в явном аналитическом виде.

Ключевые слова: надежность систем, стационарные вероятности, чувствительность, математическое и имитационное моделирование, гибридные системы передачи данных.

1. Введение

Постоянное развитие компьютерных сетей и систем передачи данных подчеркивает увеличивающуюся потребность в адекватных математических моделях и инструментальных средствах, позволяющих проводить исследование их функционирования [1]. Необходимо иметь помощь как на этапах проектирования (для сравнение решений), так и эксплуатации (управление качеством сервиса) сетевых систем. Действительно, разработка сложной технической системы требует не только качественного моделирования, чтобы проверить, насколько логически правильно она построена, но требуется также и априорная проверка надёжности и производительности системы на этапе проектирования [2].

Целью работы является проведение аналитического и имитационного моделирования дублированной системы $\langle M_2/GI/1 \rangle$ передачи данных, состоящей из двух каналов передачи данных — основного и резервного, находящегося в холодном резерве [3–5].

2. Постановка задачи и обозначения

Предположим, что элементы системы (каналы) функционируют независимо друг от друга и введем следующие обозначения:

Δ – малый промежуток времени,

A – случайная величина (с.в.) в.б.р. основного элемента,

B – с.в. времени ремонта,

$V(x)$ – функция распределения (ФР) с.в. B ,

$b(x)$ – плотность распределения (ПР) с.в. B ,

$\tilde{b}(s)$ – преобразование Лапласа (ПЛ) плотности $b(x)$,

EA – среднее в.б.р. основного элемента,

EB – среднее время ремонта отказавшего элемента,

$\rho = \frac{EA}{EB}$ – относительная скорость восстановления,

$\delta(x) = \frac{b(x)}{1-B(x)}$ – условная ПР остаточной длительности ремонта элемента, находящегося в ремонте время t ,

α – параметр экспоненциального распределения в.б.р. прибора.

2.1. Марковизация процесса

Рассмотрим случайный процесс $v(t)$ – число отказавших элементов (каналов) в момент времени t , множество состояний системы $E = \{1, 2, 3\}$.

Для описания поведения системы с помощью марковского процесса, введём дополнительную переменную $x(t) \in \mathbb{R}_+^2$ – время, затраченное в момент t , на ремонт отказавшего элемента. Получим двумерный процесс $(v(t), x(t))$, с расширенным пространством состояний $\epsilon = \{(0), (1, x), (2, x)\}$.

Обозначим $p_0(t)$ – вероятность того, что в момент времени t система находится в состоянии $i = 0$, $p_i(t, x)$ – плотность распределения (по непрерывной компоненте) вероятностей того, что в момент времени t система находится в состоянии $i (i = 1, 2)$, и время затраченное на ремонт отказавшего элемента находится в интервале $(x, x + dx)$

$$p_0(t) = p\{v(t) = 0\},$$

$$p_1(t, x)dx = p\{v(t) = 1, x < x(t) < x + dx\},$$

$$p_2(t, x)dx = p\{v(t) = 2, x < x(t) < x + dx\}.$$

С помощью формулы полной вероятности перейдем к выводу системы уравнений Колмогорова, а с предельным переходом при $\Delta \rightarrow 0$ и в предположении что процесс имеет стационарное распределение при

$t \rightarrow \infty$, получаем системы дифференциальных уравнений Колмогорова (СДУК) в виде:

$$\begin{cases} \alpha \cdot p_0 = \int_0^\infty p_1(x) \cdot \delta(x) dx \\ \frac{dp_1(x)}{dx} = -(\alpha + \delta(x)) \cdot p_1(x) \\ \frac{dp_2(x)}{dx} = -\delta(x) \cdot p_2(x) + \alpha \cdot p_1(x) \\ p_1(0) dx = \int_0^\infty p_2(x) \cdot \delta(x) dx p_0(t) \cdot \alpha \end{cases}$$

Перейдем к решению полученной СДУК, используя метод вариации постоянной, и получаем стационарные вероятности состояний системы:

$$p_0 = \frac{\tilde{b}(\alpha)}{\rho^{-1} + \tilde{b}(\alpha)}, \quad p_1 = \frac{1 - \tilde{b}(\alpha)}{\rho^{-1} + \tilde{b}(\alpha)}, \quad p_2 = \frac{\rho^{-1} - (1 - \tilde{b}(\alpha))}{\rho^{-1} + \tilde{b}(\alpha)}, \quad \text{где } \rho^{-1} = b \cdot \alpha.$$

Очевидно, имеется зависимость стационарных вероятностей состояний системы от вида распределений времени ремонта [6].

На рисунке представлены графики зависимости стационарной вероятности безотказной работы системы от относительной скорости восстановления

Видно, что эта зависимость становится исчезающе малой при «быстром» восстановлении. Явные аналитические выражения для стационарного распределения рассматриваемой системы удается получить не всегда. Поэтому возникла задача построения имитационной модели, аппроксимирующей аналитическую модель системы.

2.2. Сравнение и анализ результатов математического и имитационного моделирования

Определим следующие состояния моделируемой системы:

- состояние 0: один (главный) элемент работает, второй – в холодном резерве;
- состояние 1: один прибор отказал и находится в ремонте, второй – работает;
- состояние 2: оба прибора отказали, один – в ремонте, второй ждет своей очереди на ремонт.

Имитационное моделирование системы $\langle M_2/GI/1 \rangle$ осуществляется на основе дискретно-событийного подхода, согласно которому изменение состояния системы происходит мгновенно (в модельном времени) при наступлении того или иного события.

Для наглядности имитационная модель представлена графически на рисунке 2 в виде блок-схемы. Критерием останова основного цикла модели является достижение максимального модельного времени выполнения T .

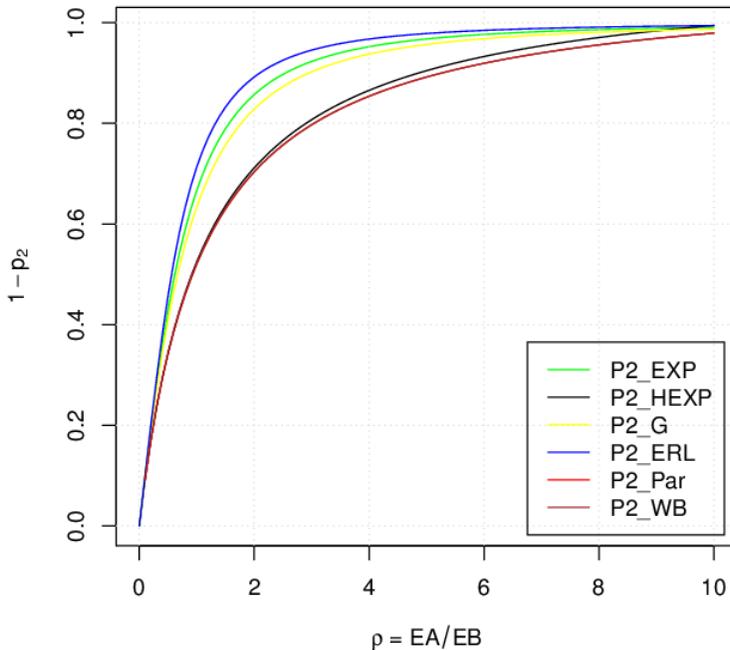


Рис. 1. Графики зависимости стационарной в.б.р. $1 - p_2$ системы от ρ для различных ФР времени ремонта

Алгоритм дискретно-событийного процесса имитационного моделирования был программно реализован на языке R. Моделирование проводилось при ограничении на максимальное модельное время $T = 100000$.

Для анализа и сравнения результатов были выбраны следующие распределения времени ремонта элементов: экспоненциальное, Эрланга, Вейбулла-Гнеденко и Парето (как пример распределения с «тяжелым хвостом»).

Из таблицы 1 видно, что результаты имитационного моделирования хорошо согласуются с результатами, полученными по явным аналитическим формулам. Также видно, что с ростом ρ различия в значениях p_i исчезают.

Для большей наглядности, проведем сравнение аналитических и имитационных результатов графически.

Из рис. 3 и рис. 4 видно, что различия между кривыми с ростом ρ становятся исчезающе малыми.

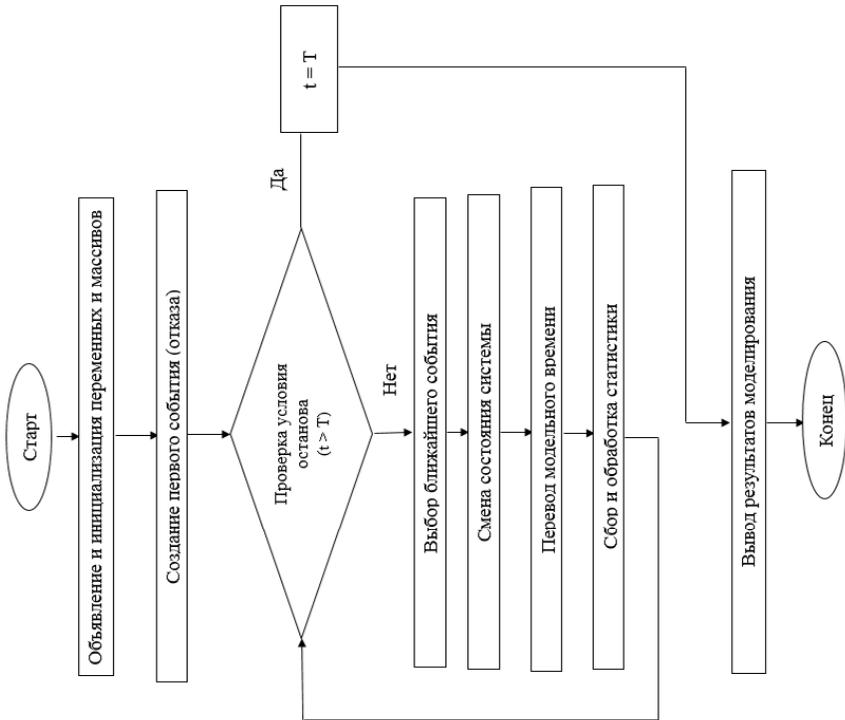


Рис. 2. Блок-схема алгоритма системы моделирования.

Построенная имитационная модель хорошо аппроксимирует аналитическую модель.

3. Заключение

В работе были получены явные аналитические выражения для стационарного распределения вероятностей состояний системы и для стационарной в.б.р. системы как в общем случае, так и для некоторых частных случаев распределений. Полученные формулы показывают наличие явной зависимости этих характеристик от вида ФР времени ремонта её элементов. Однако численные исследования и анализ построенных графиков показали, что эта зависимость становится исчезающе малой при «быстром» восстановлении, то есть с ростом относительной скорости восстановления ρ .

M_2	GI	$EXP\left(\beta = \frac{1}{EB}\right)$		$Erlang\left(\beta = \frac{k=2}{EB}\right)$		$WB\left(\beta = \frac{k=2}{EB}, \frac{1}{2}\right)$		$Pareto\left(k, \frac{k}{k-1}, \frac{1}{EB}\right)$	
		$EXP\left(\alpha = \frac{1}{EA}\right)$	p_i	Имитац.	Аналит.	Имитац.	Аналит.	Имитац.	Аналит.
$\rho = 1;$ $EA = 5$ $EB = 5$	p_0	0.33629	0.333333	0.31061	0.30769	0.37929	0.39602	0.22677	0.26985
	p_1	0.33298	0.333333	0.38505	0.38462	0.20945	0.20796	0.49506	0.46031
	p_2	0.33073	0.333333	0.30434	0.30769	0.41126	0.39602	0.27818	0.26985
$\rho = 10;$ $EA = 50$ $EB = 5$	p_0	0.91373	0.9009	0.8619	0.9007	0.91948	0.90204	0.91817	0.90049
	p_1	0.08452	0.09009	0.1252	0.09232	0.05948	0.0776	0.07749	0.09466
	p_2	0.00175	0.00901	0.01292	0.00698	0.02104	0.02036	0.00434	0.00486
$\rho = 100;$ $EA = 500$ $EB = 5$	p_0	0.9955	0.99	0.99033	0.99	0.99802	0.99	0.98522	0.99
	p_1	0.00448	0.0099	0.00967	0.00992	0.00198	0.00971	0.01478	0.00995
	p_2	0	0.0001	0	0.00007	0	0.00029	0	0.0005

Таблица 1
Стационарные вероятности p_i состояний системы $\langle M_2/GI/1 \rangle$,
рассчитанные имитационно и аналитически при разных значениях
модельного параметра $\rho = 1, 10, 100$.

Было проведено имитационное моделирование системы $\langle M_2/GI/1 \rangle$ на основе дискретно-событийного подхода. Численное и графическое сравнение результатов имитационного моделирования с аналитическими результатами показало, что, во-первых, построенная имитационная модель хорошо аппроксимирует аналитическую модель системы, а значит может быть использована в тех случаях, когда не удается получить выражения для стационарных вероятностей состояний системы в явном аналитическом виде, а во-вторых, при «быстром» восстановлении стационарные вероятности состояний системы становятся нечувствительными к виду и параметрам ФР времени ремонта элементов.

Литература

1. V.M. Vishnevsky, D.V. Kozyrev, O. V. Semenova. Redundant queuing system with unreliable servers // International Congress on Ultra Modern Telecommunications and Control Systems and Workshops. IEEE Xplore, 2015. Pp. 283 – 286.
2. Д.В. Козырев. Анализ вероятностно-временных характеристик высоконадёжных телекоммуникационных систем: диссертация ... кандидата физико-математических наук, Москва, 2013. – 128 с.: ил. РГБ ОД, 61 13-1/1005

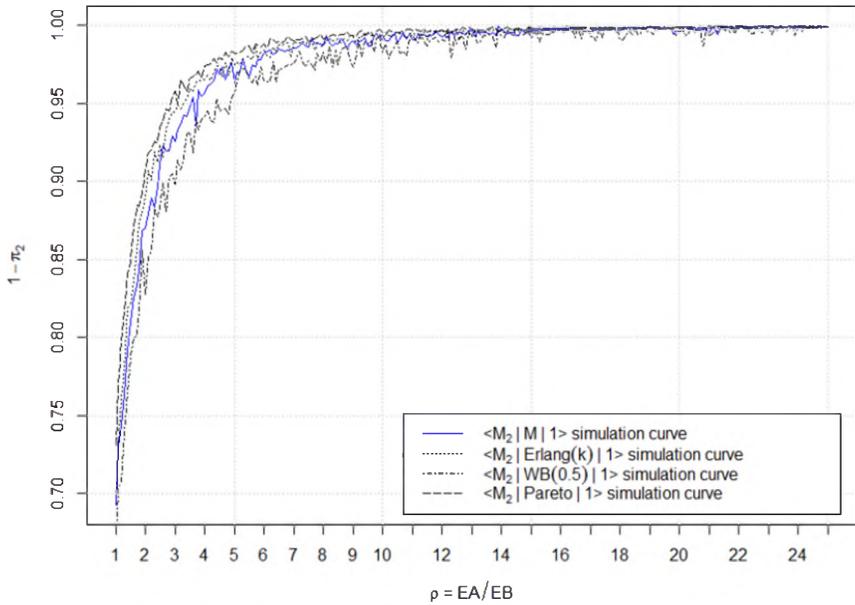


Рис. 3. Графики зависимости в.б.р. системы $1 - p_2$ от модельного параметра ρ по результатам имитационного моделирования.

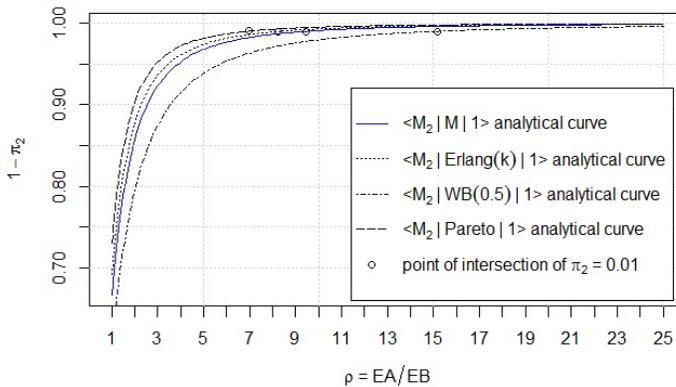


Рис. 4. Графики зависимости стационарной в.б.р. $1 - p_2$ системы от ρ по результатам аналитического моделирования

3. Б.В. Гнеденко. О ненагруженном дублировании // Изв. АН СССР. Тех. кибернетика. 1964. № 4. С. 3–12.
4. Б.В. Гнеденко. О дублировании с восстановлением // Изв. АН СССР. Тех. кибернетика. 1964. № 5. С. 111–118.
5. А.Д. Соловьёв. Резервирование с быстрым восстановлением // Изв. АН СССР. Тех. кибернетика. 1970. № 1. С. 56–71.
6. В. Рыков, Чан Ань Нгиа. О чувствительности характеристик надежности систем к виду функций распределения времени безотказной работы и восстановления их элементов. Вестник Российского университета дружбы народов. Серия: Математика. Информатика. Физика. № 3 (2014), С. 65–77.

UDC 519.217

Sensitivity analysis of steady state reliability characteristics of a cold redundant data transmission system to the shapes of lifetime and repair time distributions of its elements

Н. Г. К. Houankpo*, D. V. Kozyrev*[†]

* *Department of Applied Probability and Informatics, RUDN University, Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

[†] *V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Profsoyuznaya str. 65, Moscow, Russia*

We consider the problem of sensitivity analysis of a redundant system's reliability characteristics to shapes of the input distributions. Discrete-event simulation model is introduced to obtain results in case of a general (non-exponential) distribution of repair time of the elements.

Keywords: system reliability, steady state probabilities, sensitivity, mathematical modeling and simulation, hybrid data transmission systems.

UDC 004.627

Analysis of background modeling methods performance in lossy video compression systems

N. Ustyuzhanin*[†], M. Gilmutdinov*

* *Saint Petersburg State University of Aerospace Instrumentation (SUAI), 67, Bolshaya Morskaya str., Saint-Petersburg, 190000, Russia*

[†] *Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow, 143026, Russia*

Abstract.

Video streams is a significant part of Internet traffic and many cloud storages. One of specific types of video stream is generated by video surveillance systems. Video surveillance traffic increases extremely per year so it should be compressed efficiently. This type of traffic has several features specific for this application. One of them is background scene which is usually static or it changes in time insignificantly. Methods of background modeling is highly developed and well studied in computer vision applications but they are not used explicitly in modern compression standards, e.g. ITU-T h.264/AVC or ITU-T h.265 (High efficient video coding - HEVC). Using such type of video processing in lossy compression algorithms meets a problem of quantization noise affect. This paper presents the results of analysis of quantization noise impact on background modeling performance. State-of-the-art background modeling methods are considered on open sequences with video surveillance specifics. The input is data recovered after ITU-T h.264/AVC or HEVC compression with various rate-distortion factors. Background modeling performance is estimated using ground truth data and Peak-Signal-to-Noise-Ratio (PSNR) criterion.

Keywords: video surveillance, computer vision, background modeling, lossy video compression standards, quantization noise affect.

1. Introduction

Using background modeling methods is one of the general way to improve compression rate in modern video surveillance systems. Background modeling algorithms are used for moving objects detection in computer vision. This approaches can be successfully applied in video compression because several assumptions of typical video surveillance system are valid [1]. Main of them are fixed position of camera view, static or slight changing background scene, fixed values of camera parameters, e.g. focus, exposure, white-balance. Unfortunately not all of these assumptions are true in practice. Additional issue which appears in video compression systems is quantization noise that appears after lossy compression. This work is dedicated to preliminary analysis of background modeling feasibility in video compression algorithms applied to video surveillance system content. This type of computer vision and video compression algorithms synergy is not yet studied enough. Chinese national video compression standard AVS also known as IEEE-1857 [2] uses background frames for video surveillance

content. But they have limited methods of background estimation applied at encoder side only. It is very important to analyze quantization noise influence on various background modeling algorithms performance because it might have an impact on modeling process especially at the decoder side.

ITU-T h.264/AVC and ITU-T h.265/HEVC decoders output is used in analysis as a background modeling input. Open video sequences with ground truth information are used to estimate background modeling accuracy. Background modeling performance is evaluated via Peak-Signal-to-Noise-Ratio (PSNR) calculated between ground truth data and background estimations which use video decoder output.

The remaining part of this paper is organized as follows. Next section contains a description of state-of-the-art background modeling algorithm considered in this investigation. Performance evaluation framework used for background modeling analysis is described in section 3. Section 4 presents experimental result analysis of quantization noise impact.

2. State-of-the-art background modeling methods

Let us consider a typical background estimation algorithm. It consists of two main steps.

1. *Initialization or training* that can use first \mathcal{I} frames.
2. *Maintenance* updates background information. At least currently processed frame should be used for this update but it can also include frames previously processed .

Simplest background model, e.g. Background Modeling Independent of Time described in [3], assumes that the first frame in video sequence is completely background and it does not change in time. Thus initialization step is

$$B_{x,y}^0 = I_{x,y}^0,$$

where $B_{x,y}^{(t)}$ is background pixel value at position (x, y) in frame t . This model does not require update step, so

$$B_{x,y}^{t+1} = B_{x,y}^t,$$

Maintenance step of another simple estimation method uses thresholded comparison of difference of pixels at similar position in adjacent frames.

A class of simple background estimation methods uses moving calculations of \mathcal{K} previous frames. It can be *moving average (Av)*:

$$B_{x,y}^t = \frac{1}{\mathcal{K}} \sum_{i=1}^{\mathcal{K}} I_{x,y}^{t-i}$$

or *moving median (Med)*. Moving median result is element with index $\lfloor \frac{K}{2} \rfloor$ of sorted set $\{I_{x,y}^{t-K}, \dots, I_{x,y}^{t-1}\}$. Sorting order is not significant. To reduce a high amount of memory required for methods from this class, *Running Gaussian Averaging (RGA)* Background Modeling [3], [4] uses infinite impulse response filter to simplify averaging procedure. Its initialization and maintenance steps can be expressed as:

$$B_{x,y}^t = \begin{cases} I_{x,y}^0 = 0, & t = 0 \\ (1 - \alpha)B_{x,y}^{t-1} + \alpha I_{x,y}^t, & t > 0 \end{cases},$$

where $\alpha \in [0, 1]$ is called *learning rate*. Typically it is set to a small number as 0.05 .

To solve the problems with extreme lighting and background changing, is possible to use a probabilistic models. One of them is *Mixture of Gaussians (MoG)* [5], [3], [1]. MoG helps to adapt to environment changes, such as camera automatic adjustments, camera jitter, moved background object moving, lightning changes, etc. This model uses several assumptions:

Assumption 1: For a single component frames each pixel intensity distribution is considered as a mixture of several random processes:

$$P(I_{x,y}^t) = \sum_{i=1}^K \omega_i^t \cdot \eta(I_{x,y}^t, m_i^t, (\sigma_i^t)^2),$$

where K is total number of processes.

Assumption 2: at the moment t each process is described by a normal distribution $\eta \sim N(m_i^t, (\sigma_i^t)^2)$ with expectation m_i^t , variance $(\sigma_i^t)^2$ and weight ω_i^t

Assumption 3: sum of all weights is equal to one:

$$\sum_{i=1}^K \omega_i^t = 1$$

Lobster method [6], [1] is designed for moving objects detection in frequently changing luminance. It uses *LBSP* descriptor previously introduced in [7]. *LBSP* descriptor is described by the following equations:

$$LBSP(x, y) = \sum_{p=0}^{P-1} d_{T_r}(I_{x+\delta x, y+\delta y}^t, I_{x,y}^t) \cdot 2^p$$

with

$$d_{T_r}(I_{c+\delta}, I_c) = \begin{cases} 1, & \text{if } |I_{c+\delta} - I_c| \leq T_r \cdot I_c \\ 0, & \text{otherwise} \end{cases}$$

where T_r is threshold value, $I_{x,y}^t$ corresponds to the intensity of the central (analyzed) pixel, and $I_{x+\delta x,y+\delta y}^t$ corresponds to the intensity of the neighbor at position $(x + \delta x, y + \delta y)$.

Different background modeling methods generate different artifacts. Figure 1 depicts example of background frames obtained by the different methods in the sequence with moving pedestrians.



Figure 1. Typical visual artefacts in background models: a) running gaussian averaging, b) median, c) average, d) mixture of gaussians

3. Performance evaluation framework

It is assumed that the background frames are generated at the encoder and decoder by a similar algorithm (so-called decoder-based approach). Typical video codec uses *group of pictures* (GOP) structure. It is a sub-sequence of frames started for intra-encoded frames. This first frame is usually called Instantaneous Decoder Refresh (IDR) frame because all previous information resets at the decoder side to default settings. First N_{BG} frames of GOP are used for background initialization training. Starting from frame $(N_{BG} + 1)$ background frame can be used as reference information both at the encoder and decoder sides. Background frame requires update after each frame processing. The time-line of this algorithm is depicted in fig. 2.

The main scheme of proposed framework for evaluation of quantization noise impact is depicted in fig. 3. To provide input data for background modeling initial video sequence is encoded and then decoded to obtain reconstructed video frames. After background modeling background frames are compared with ground truth data using *Peak Signal-To-Noise* (PSNR) criterion. To estimate affect of quantization noise with different parameters, encoder uses various quantization steps.

4. Experimental results

4.1. Experiment setup

Two modern video codecs are used in experiment described in section 3:

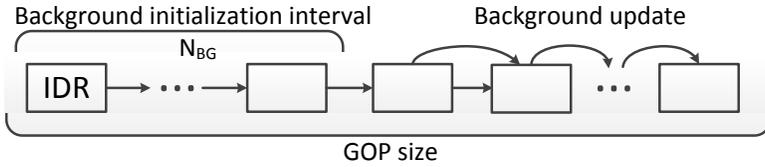


Figure 2. Background modeling process

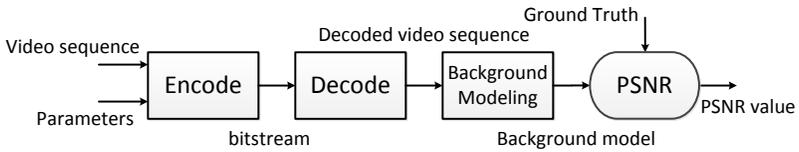


Figure 3. Scheme of experiment

- x264 [8] is implementation of ITU-T H.264/AVC standard [9]. Baseline profile is used, because it is appropriate for real-time video application.
- x265 [10] is implementation of ITU-T H.265/HEVC standard [11]. x265 uses main profile without B-frames.

Both codecs use 20 frames IDR-frame period. Frame rate is 10 frames per second. That is enough for video surveillance systems.

Two open video sequences with ground truth frames are used in experiment:

- 111 dataset [12],
- 121 dataset [13].

These sequences are synthetic. It is need to avoid camera noise affect. All pictures of both sequences have 640×480 resolution. They are initially presented in PNG lossless format. Examples of round truth frame and frame of the sequences are depicted in fig. 4.

For experiment each sequence was encoded with different quantization steps described via *Constant Rate Factor (CRF)* codecs parameter. It varies from 0 (lossless) to 51 (coarsest quality). Background frames for each sequences are generated using methods mostly described in section 2. Besides proprietary implementation some of methods (Lobster, Advanced Background Learning (ABL), Mixture of Gaussians (MoG)) are used from open source BGS Library [14], [15]. Moving average (Av) and moving

median (Med) methods also included to comparison to estimate their dependence on quantization noise.



Figure 4. a) ground truth, b) t -th frame (111 dataset [12])

4.2. Experimental results analysis

Curves of $PSNR/CRF$ are shown in fig. 5. Performed experiments allow to determine some of background modeling methods which are sensitive to quantization noise significantly. There are Running Gaussian Averaging (RGA) and Lobster. Other methods have quiet similar performance especially for the medium and low quality compression parameters. It should be noted that moving average (Av) and MoG by Stauffer and Grimson version [5] have a slight lower performance in case of high quality compression for 111 dataset. Performance of most methods decreases monotonically while compression ratio increases. Moving median (Med) should be excluded from list of appropriate method in spite of acceptable results. The main reason is high memory requirement.

5. Conclusion

Background modeling is promising approach that can improve compression performance in video surveillance systems. To avoid background frame transmission it can be generated similarly both at the encoder and the decoder sides (decoder-based approach). This paper is dedicated to analysis of quantization noise impact on background modeling performance in case of decoder-based approach. Several methods mostly used for background modeling in computer vision applications were analyzed. Open video sequences with ground truth background data is used for analysis. According to this analysis MoG-based approaches can be considered as most insensitive to quantization noise. They show the best results for sequences, processed by modern video compression standards ITU-T h.264 and ITU-T h.265.

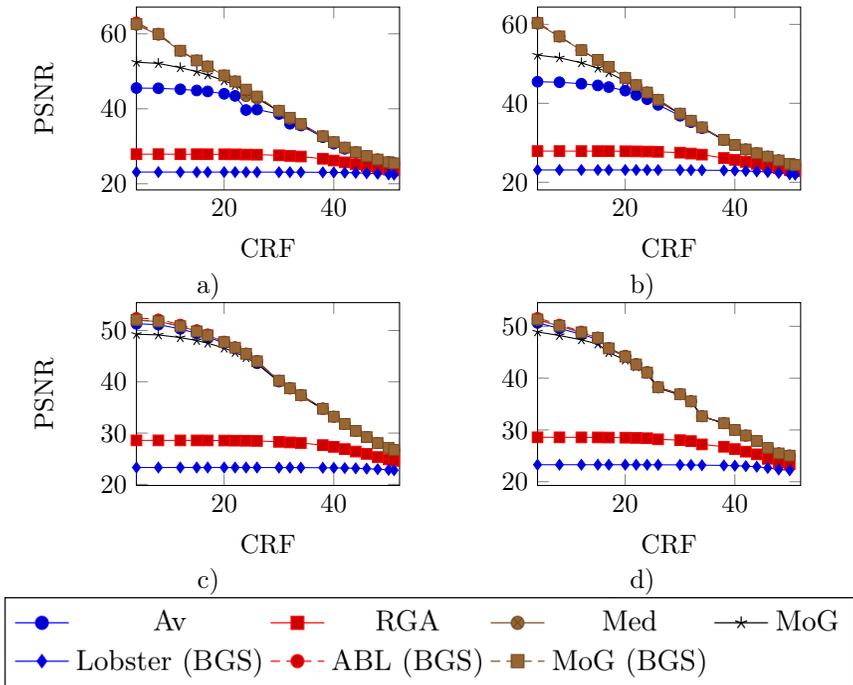


Figure 5. Diagram of dependencies of PSNR from CRF for video sequences: a) 111 [12] (x265 [10]), b) 111 [12] (x264 [8]), c) 121 [13] (x265 [10]), d) 121 [13] (x264 [8])

References

1. Bouwmans T., Porikli F., Hoferlin B., Vacavant A. // Background Modeling and Foreground Detection for Video Surveillance. — Taylor and Francis Group, LLC, 2015.
2. Gao W., Ma S. Advanced video coding systems. — Springer, 2014. — P. 244.
3. Hassanpour H., Sedighi M., Manashty A. R. Video frames background modeling: Reviewing the Techniques. — 2011. — P. 72–78.
4. Everette S. Gardner J. Exponential smoothing: The state of the art Part II // International Journal of Forecasting. — 30 June 2006. — P. 637–666.
5. Stauffer C., Grimson W. Adaptive background mixture models for real - time tracking // Proc IEEE Conf on Comp Vision and Patt Recog

- (CVPR 1999) / IEEE. — 1999. — P. 246–252.
6. St-Charles P.-L., Bilodeau G.-A. Improving background subtraction using local binary similarity patterns // Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. — 23 June 2014. — P. 509–515.
 7. Bilodeau G.-A., Jodoin J.-P., Saunier N. Change detection in feature space using local binary similarity patterns // In Computer and Robot Vision. — 2013. — P. 106–112.
 8. x264 source code. — <http://www.videolan.org/developers/x264.html>. — [Online; accessed 30-September-2016].
 9. Wiegand T., Sullivan G. J. Overview of the h.264/avc video coding standard // IEEE Transactions on circuits and systems for video technology, vol. 13, no. 7. — july 2003.
 10. x265 source code. — <http://www.videolan.org/developers/x265.html>. — [Online; accessed 30-September-2016].
 11. High efficiency video coding // Recommendation ITU-T H.265. — 04/2015.
 12. 111 dataset. — http://bmc.iut-auvergne.com/wp-content/ressources/videos/learning/111_png.zip. — [Online; accessed 30-September-2016].
 13. 121 dataset. — http://bmc.iut-auvergne.com/wp-content/ressources/videos/learning/121_png.zip. — [Online; accessed 30-September-2016].
 14. Sobral A., Bouwmans T. Bgs library: A Library Framework for Algorithms Evaluation in Foreground/Background Segmentation // Background Modeling and Foreground Detection for Video Surveillance. — CRC Press, Taylor and Francis Group., 2014.
 15. Sobral, Andrews. BGSLibrary: An OpenCV C++ Background Subtraction Library // IX Workshop de Visio Computacional (WVC'2013). — Rio de Janeiro, Brazil, 2013. — Jun. — URL: <https://github.com/andrewssobral/bgslibrary>.

UDC 519.218.31

Stationary waiting time in $G/M/n/r$ FCFS queue with random renovation

I.S. Zaryadov^{*†}, R.V. Razumchik^{†*}

** Department of Applied Probability and Informatics,
Peoples' Friendship University of Russia (RUDN University),
6 Miklukho-Maklaya str., Moscow, 117198, Russia*

*† Institute of Informatics Problems, FRC CSC RAS,
44-2 Vavilova Str., Moscow 119333, Russia*

Abstract. Consideration is given to $G/M/n/r$ FCFS queue with random renovation policy. This policy implies that a customer upon service completion with probability $q(l)$ removes l customers from the queue and then leaves the system. The choice of customers to be removed from the queue is done in a random fashion. Customer from the queue are served in FCFS order. For this system in a series of papers there were obtained stationary characteristics related to the number of customers in the system. Here we obtain several expressions for the stationary time-related characteristics and specifically dwell on the stationary waiting time distribution, which is obtained in terms of the transforms.

Keywords: random renovation, finite capacity, stationary characteristics, waiting time.

1. Introduction

Queueing systems with losses are frequently encountered in telecommunication networks modelling. Losses usually happen due to the limited buffer capacity but there are also cases, when a customer accepted to the buffer is lost before service — due to negative customers arrival [1, 2], servers failures or due to special service disciplines (for example, general renovation [3–9]).

This paper is structured as follows. In Section 2 the description of the queueing system is given. Section 3 is devoted to auxiliary probabilities for time-related characteristics. In Sections 4 and 5 the stationary waiting time distribution of the «killed» and served customer is considered.

In order to obtain some characteristics for our system we will use the standard approaches from [10].

2. System description

Consideration is given to the $G/M/n/r$ queueing system. Times between arrivals are i.i.d. with the distribution function $A(x)$ and finite mean $a = \int_0^\infty x dA(x)$. Denote by $\bar{A}(x) = 1 - A(x)$. Service times are exponentially distributed with rate μ . An arriving customer which finds system full, is lost. Customers are served from the queue in FCFS order.

It is assumed that random renovation policy is implemented in the system. This policy implies that a customer upon service completion removes («kills») l customers from the queue with probability $q(l)$, if there are more than l customers in the queue and with the probability $Q(l) = \sum_{k=l}^r q(k)$ empties the queue. The probabilities $q(l)$, $0 \leq l \leq r - 1$, and $Q(r) = q(r)$ according to [6] are called renovation probabilities. As usual the normalization condition $Q(0) = \sum_{l=0}^r q(l) = 1$ must hold. Clearly $q(0) = p$ is the probability that the customer upon service completion leaves the system without removing any customers from the queue. The choice of positions in the queue from which the customers have to be removed is done in a purely random fashion i.e. all outcomes are equally likely.

In order to compute the stationary waiting time distribution one needs several results from [7]: the stationary distribution p_i^- , $i = 0, n + r - 1$, of the number of customers in the system *at arrival instants*, stationary loss probability π , stationary probability p_{loss} that an arbitrary customer will be «killed» from the queue. Notice that these quantities do not depend on the service discipline and the order in which the customers are removed from the queue.

The process $\xi(t) = \{(k(t), x(t)), t \geq 0\}$, where $N(t)$ denotes the number of customers in the system at instant t , and $X(t)$ – is the elapsed time since the last arrival, is the linearwise Markov process.

It is known that any linearwise process with the finite state set of the discrete component (which is $N(t)$) is ergodic (see [10, Section 1.6]).

3. Auxiliary quantities

Before one can proceed to the calculation of the stationary waiting time distributions, several preliminary results are needed.

Let

$A_{(0,x]}$ denote the number of arrivals during the interval $(0, x]$;

$S_{(0,x]}$ denote the number of service completions during the interval $(0, x]$;

$F(t)$ denote the number of customers *in front* of the tagged customer at instant t ;

$B(t)$ denote the number of customers *behind* the tagged customer at instant t .

Assume that at some time instant τ the queue is not empty and choose any customer in the queue. Let us call it a tagged customer. It can happen that, once the tagged customer is chosen, a number of other customers may be in front (excluding those in service) and behind it in the queue.

Define the following probabilities:

$$\begin{aligned} \pi_m(i, j; i', j'; x) = \\ = \mathbf{P} \{F(x) = i', B(x) = j' | F(0) = i, B(0) = j, A_{(0,x]} = 0, S_{(0,x]} = m\}, \end{aligned}$$

$$m = \overline{1, r - 1}, i = \overline{m, r - 1}, j = \overline{0, r - i - 1}, i' = \overline{0, i - m}, j' = \overline{0, j},$$

$$\tilde{\pi}_m(i, j) = \mathbf{P} \{ \text{customer has entered service after } m\text{-th service completion} | F(0) = m, \\ m = \overline{1, r - 1}, i = \overline{m - 1, r - 1}, j = \overline{0, r - i - 1}. \}$$

Assume at some instant, say τ , there is a customer in the queue arrives at the system, finds $1 \leq m \leq r - 1$ customers in the queue. Denote by $F(\tau+)$ the number of customers in front of the tagged customer in the queue.

Let $A(x)$ and $N(x)$ denote the number of arrivals and service completions that occur during the interval $(0, x]$.

Denote by $F(t)$ the number of customers in front of the tagged customer at time t ; by $B(t)$ denote the number of customers behind the tagged customer at time t . Notice that due to the random renovation policy after each service completion, the tagged customer may be either «killed», or enter service or remain in the queue. Let S_m be the status of the tagged customer immediately after m -th service completion ($S_m = 0$ means that the tagged customer is in the queue immediately after the m -th service completion; $S_m = 1$ means that the tagged customer is in the server immediately after the m -th service completion; $S_m = -1$ means that the tagged customer was «killed» by the m -th served customer). If $S_m = -1, S_n, n > m$ are not defined.

Without the loss of generality in the steady state one can put $t = 0$.

Define the following auxiliary quantities:

$$\pi_m(i, j; i', j') = \\ = \mathbf{P} \{ F(x) = i', B(x) = j' | F(0) = i, B(0) = j, A(x) = 0, N(x) = m \}, \\ m = \overline{1, r - 1}, i = \overline{m, r - 1}, j = \overline{0, r - i - 1}, i' = \overline{0, i - m}, j' = \overline{0, j},$$

$$\tilde{\pi}_m(i, j) = \mathbf{P} \{ S_m = -1 | F(0) = i, B(0) = j, A(x) = 0, N(x) = m \}, \\ m = \overline{1, r - 1}, i = \overline{m - 1, r - 1}, j = \overline{0, r - i - 1}.$$

For example, $\pi_m(i, j; i', j')$ is the probability that at time instant x there will be i' customers in front and j' customers behind the tagged customer, provided that at time 0 there were i customers in front and j customers behind the tagged customer, in the interval $(0, x]$ one had $m, m = \overline{1, r - 1}$, service completions and no new arrivals.

The relations for $\pi_m(i, j; i', j')$ and $\tilde{\pi}_m(i, j)$ can be written out using the first step analysis. For $\pi_m(i, j; i', j')$ one has:

$$\pi_m(i, j; i', j') = \sum_{i''=i'+1}^{i-m+1} \sum_{j''=j'}^j \pi_{m-1}(i, j; i'', j'') \pi_1(i'', j''; i', j'),$$

$$m = \overline{2, r - 1}, \quad i = \overline{m, r - 1}, \quad j = \overline{0, r - i - 1}, \quad i' = \overline{0, i - m}, \quad j' = \overline{0, j}, \quad (1)$$

where

$$\pi_1(i, j; i', j') = q(i - i' - 1 + j - j') \frac{C_i^{i'+1} C_j^{j'}}{C_{i+j+1}^{i'+j'+2}},$$

$$i = \overline{1, r - 1}, \quad j = \overline{0, r - i - 1}, \quad i' = \overline{0, i - 1}, \quad j' = \overline{0, j}.$$

Here and henceforth C_n^k denotes the number of k -combinations from a set of n elements.

For the probability $\tilde{\pi}_m(i, j)$ the following relations hold:

$$\tilde{\pi}_m(i, j) = \sum_{i'=0}^{i-m+1} \sum_{j'=0}^j \pi_{m-1}(i, j; i', j') \tilde{\pi}_1(i', j'),$$

$$m = \overline{2, r - 1}, \quad i = \overline{m - 1, r - 1}, \quad j = \overline{0, r - i - 1},$$

where

$$\tilde{\pi}_1(i', j') = \sum_{i''=1}^{i'+j'} q(i'') \frac{i''}{i' + j' + 1} + Q(i' + j' + 1),$$

$$i' = \overline{0, r - 1}, \quad j' = \overline{0, r - 1 - i'}. \quad (2)$$

4. Stationary waiting time of the «killed» customer

We will start with the waiting time of the «killed» customer, i.e. the customer, which has entered the system but was «killed» (was dropped from the buffer) while waiting for service in the queue. Let

$$W_{i,j}^{(\text{loss})}(x) = \mathbf{P} \{T < x, \cup_{m=1}^{i+n} S_m = -1 | F(0) = i, B(0) = j\},$$

$$i = \overline{0, r - 1}, \quad j = \overline{0, r - 1 - i},$$

where $T -$ is the tagged customer's waiting time. The CDF $W_{i,j}^{(\text{loss})}(x)$ is the probability that the tagged customer will be «killed» in the time interval $(0, x)$, given that initially there were i customers in front (excluding those in service) and j customers behind it. Then the waiting time CDF of the arriving accepted customer which finds $j + n$ customers in the system and which will be «killed» while waiting is $W_{0,j}^{(\text{loss})}(x)$.

In order to calculate $W_{i,j}^{(\text{loss})}(x)$, one needs the following auxiliary quantities:

$$\tilde{U}_{k,i}(x) = \mathbf{P} \{T < x, \cup_{m=1}^{i+n} S_m = -1 | F(0) = i, B(0) = k - i, A(x) = 0\},$$

$$i = \overline{0, r - 1}, \quad k = \overline{0, r},$$

$$\begin{aligned} & u_{m, (k,l)}^{(i,j)}(x) = \\ & = \mathbf{P} \{ F(x) = i - k, B(x) = j - l, N(x) = m | F(0) = i, B(0) = j, A(x) = 0 \}, \\ & \quad m = \overline{1, r - 1}, \quad i = \overline{m - 1, r - 1}, \quad j = \overline{0, r - i - 1}, \quad k = \overline{0, i}, \quad l = \overline{0, j}. \end{aligned}$$

Denote by $w_{i,j}^{(\text{loss})}(x)$ and $\tilde{u}_{k,i}(x)$ the probability density functions of the CDF $W_{i,j}^{(\text{loss})}(x)$ and $\tilde{U}_{k,i}(x)$.

Using the law of total probability and (1)–(2), one can write out the exact expressions for the $\tilde{u}_{k,i}(x)$, $u_{m, (k,l)}^{(i,j)}(x)$. We have

$$u_{m, (k,l)}^{(i,j)}(x) = \pi_m(i, j; i - m - k, j - l) \frac{(\mu n x)^m}{m!} e^{-\mu n x},$$

$$m = \overline{1, r - 1}, \quad k = \overline{0, i}, \quad l = \overline{0, j}, \quad i = \overline{m - 1, r - 1}, \quad j = \overline{0, r - 1 - i},$$

$$u_{0, (0,0)}^{(i,j)}(x) = u_0(x) = e^{-\mu n x},$$

$$u_{0, (k,l)}^{(i,j)}(x) = 0,$$

$$\tilde{u}_{k,i}(x) = \sum_{m=1}^{i+1} \tilde{\pi}_m(i, k - i - 1) \frac{(\mu n)^m x^{m-1}}{(m-1)!} e^{-\mu n x}, \quad k = \overline{0, r}, \quad i = \overline{0, r - 1}.$$

Denote by $\alpha(s)$ and $\tilde{\alpha}(s)$ the Laplace-Stieltjes transformation (LST) and Laplace transformation (LT) of the CDF $A(x)$, and by $\alpha^{(m)}(s)$ and $\tilde{\alpha}^{(m)}(s)$ – m -th derivatives of $\alpha(s)$ and $\tilde{\alpha}(s)$ with respect to s .

As we will obtain the waiting time in terms of LST, it is convenient now to switch from functions $\tilde{u}_{k,i}(x)$ and $u_{m, (k,l)}^{(i,j)}(x)$ to the transforms $\tilde{u}_{k,i}(s)$

and $u_{m, (k,l)}^{(i,j)}(s)$, which are defined as follows:

$$u_{m, (k,l)}^{(i,j)}(s) = \int_0^\infty e^{-sx} u_{m, (k,l)}^{(i,j)}(x) dA(x) =$$

$$= (-1)^m \pi_m(i, j; i - m - k, j - l) \frac{(\mu n)^m}{m!} \alpha^{(m)}(\mu n + s),$$

$$m = \overline{1, r - 1}, \quad k = \overline{0, i}, \quad l = \overline{0, j}, \quad i = \overline{m - 1, r - 1}, \quad j = \overline{0, r - 1 - i}, \quad (3)$$

$$u_0(s) = \alpha(\mu n + s),$$

$$\begin{aligned} \tilde{u}_{k,i}(s) &= \int_0^{\infty} \tilde{u}_{k,i}(x) \bar{A}(x) dx = \\ &= \sum_{m=1}^{i+1} \tilde{\pi}_m(i, k-i-1) \left(\left(\frac{\mu n}{\mu n + s} \right)^m + (-1)^m \frac{(\mu n)^m}{(m-1)!} \tilde{\alpha}^{(m-1)}(\mu n + s) \right), \\ & \quad k = \overline{0, r}, \quad i = \overline{0, r-1}. \quad (4) \end{aligned}$$

Now, using the first step analysis we can write out the system of equations for $w_{i,j}^{(\text{loss})}(x)$:

$$w_{0,j}^{(\text{loss})}(x) = \tilde{u}_{j+1,0}(x) \bar{A}(x) + \int_0^x u_0(y) dA(y) w_{0,j+1}^{(\text{loss})}(x-y), \quad j = \overline{0, r-2}, \quad (5)$$

$$w_{0,r-1}^{(\text{loss})}(x) = \tilde{u}_{r,0}(x) \bar{A}(x) + \int_0^x u_0(y) dA(y) w_{0,r-1}^{(\text{loss})}(x-y). \quad (6)$$

$$\begin{aligned} w_{i,j}^{(\text{loss})}(x) &= \tilde{u}_{i+j+1,i}(x) \bar{A}(x) + \\ &+ \sum_{m=1}^i \sum_{k=0}^{i-m} \sum_{l=0}^j \int_0^x u_{m,(i,j)}^{(k,l)}(x) dA(y) w_{i-m-k,j-l+1}^{(\text{loss})}(x-y) + \\ &+ \int_0^x u_0(x) dA(y) w_{i,j+1}^{(\text{loss})}(x-y), \quad i = \overline{1, r-2}, \quad j = \overline{0, r-i-2}. \quad (7) \end{aligned}$$

$$\begin{aligned} w_{i,r-i-1}^{(\text{loss})}(x) &= \tilde{u}_{r,i}(x) \bar{A}(x) + \\ &+ \sum_{m=1}^i \sum_{k=0}^{i-m} \sum_{l=0}^j \int_0^x u_{m,(i,r-i-1)}^{(k,l)}(x) dA(y) w_{i-m-k,r-i-l}^{(\text{loss})}(x-y) + \\ &+ \int_0^x u_0(y) dA(y) w_{i,r-i-1}^{(\text{loss})}(x-y). \quad i = \overline{1, r-2}, \quad (8) \end{aligned}$$

$$w_{r-1,0}^{(\text{loss})}(x) = \tilde{u}_{r,r-1}(x) \bar{A}(x) +$$

$$\begin{aligned}
 & + \sum_{m=1}^i \sum_{k=0}^{r-1-m} \int_0^x u_{m, \binom{r-1,0}{(k,0)}}(x) dA(y) w_{r-1-m-k,1}^{(\text{loss})}(x-y) + \\
 & \qquad \qquad \qquad + \int_0^x u_0(y) dA(y) w_{r-1,0}^{(\text{loss})}(x-y). \quad (9)
 \end{aligned}$$

The system of equations for LST of (5)–(6) (using (3)–(4)):

$$\omega_{0,j}^{(\text{loss})}(s) = \tilde{u}_{j+1,0}(s) + \omega_{0,j+1}^{(\text{loss})}(s) u_0(s), \quad j = \overline{0, r-2},$$

and

$$\omega_{0,r-1}^{(\text{loss})}(s) = \tilde{u}_{r,0}(s) + u_0(s) \omega_{0,r-1}^{(\text{loss})}(s)$$

The solution of this system has the form:

$$\omega_{0,j}^{(\text{loss})}(s) = \sum_{m=j+1}^{r-1} \tilde{u}_{m,0}(s) [u_0(s)]^{m-j-1} + \tilde{u}_{r,0}(s) \frac{[u_0(s)]^{r-j-1}}{1-u_0(s)}, \quad j = \overline{0, r-1}.$$

The relations (7)–(8) for the $w_{i,j}^{(\text{loss})}(x)$, $i = \overline{1, r-2}$, $j = \overline{0, r-i-2}$, in terms of LST have the form:

$$\begin{aligned}
 \omega_{i,j}^{(\text{loss})}(s) & = \tilde{u}_{i+j+1,i}(s) + \sum_{m=1}^i \sum_{k=0}^{i-m} \sum_{l=0}^j u_{m, \binom{i,j}{(k,l)}}(s) \omega_{i-m-k,j-l+1}^{(\text{loss})}(s) + \\
 & + u_0(s) \omega_{i,j+1}^{(\text{loss})}(s), \quad i = \overline{1, r-2}, \quad j = \overline{0, r-i-2}.
 \end{aligned}$$

For the LST of $w_{i,r-i-1}^{(\text{loss})}(x)$ one has

$$\begin{aligned}
 \omega_{i,r-i-1}^{(\text{loss})}(s) & = \tilde{u}_{r,i}(s) + \sum_{m=1}^i \sum_{k=0}^{i-m} \sum_{l=0}^{r-i-1} u_{m, \binom{i,r-i-1}{(k,l)}}(s) \omega_{i-m-k,r-i-l}^{(\text{loss})}(s) + \\
 & + u_0(s) \omega_{i,r-i-1}^{(\text{loss})}(s). \quad i = \overline{1, r-2}.
 \end{aligned}$$

For the fixed i this is the system of linear equations which can be solved recursively. Its solution $\omega_{i,j}^{(\text{loss})}(s)$, $i = \overline{1, r-2}$, $j = \overline{0, r-i-1}$ can be computed by

$$\begin{aligned}
 \omega_{i,j}^{(\text{loss})}(s) & = \sum_{m=j+i+1}^{r-1} \tilde{u}_{m,i}(s) [u_0(s)]^{m-j-i-1} + \tilde{u}_{r,i}(s) \frac{[u_0(s)]^{r-j-i-1}}{1-u_0(s)}, \\
 & \qquad \qquad \qquad j = \overline{0, r-1-i},
 \end{aligned}$$

$$\bar{u}_{t,i}(s) = \tilde{u}_{t,i}(s) + \sum_{m=1}^i \sum_{k=0}^{i-m} \sum_{l=0}^{t-i-1} u_{m,(i,t-i-1)}(s) \omega_{i-m-k,t-i-l}^{(\text{loss})}(s),$$

$$t = \overline{i + j + 1}, r.$$

For the fixed i , starting from $i = 1$, the computation must be performed for j from $r - 1 - i$ down to 0.

The last unknown LST of $w_{r-1,0}^{(\text{loss})}(x)$ can be computed from the equation

$$\omega_{r-1,0}^{(\text{loss})}(s) = \tilde{u}_{r,r-1}(s) + \sum_{m=1}^i \sum_{k=0}^{r-1-m} u_{m,(r-1,0)}(s) \omega_{r-1-m-k,1}^{(\text{loss})}(s) + u_0(s) \omega_{r-1,0}^{(\text{loss})}(s),$$

which follows from (9). Its solution is

$$\omega_{r-1,0}^{(\text{loss})}(s) = \frac{1}{1 - u_0(s)} \times \left(\tilde{u}_{r,r-1}(s) \sum_{m=1}^i \sum_{k=0}^{r-1-m} u_{m,(r-1,0)}(s) \omega_{r-1-m-k,1}^{(\text{loss})}(s) \right).$$

The moments of the waiting time follow immediately from the properties of the LST. For example, the mean waiting time of the accepted customer, which was «killed» while waiting for service, is

$$\bar{\omega}^{(\text{loss})} = \frac{1}{(1 - \pi)p_{\text{loss}}} \sum_{i=0}^{r-1} \bar{\omega}_i^{(\text{loss})} p_{n+i}^-,$$

where $\bar{\omega}_i^{(\text{loss})} = -(\omega_{i,0}^{(\text{loss})}(s))' \Big|_{s=0}$ and p_{loss} – is the stationary probability that a customer is «killed» (it was obtained in [7]).

5. Stationary waiting time of the served customer

Using the same methodology as in the previous section one can obtain the stationary waiting time distribution (in LST form) for served customer.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (grants 15-07-03007, 15-07-03406).

References

1. *Gelenbe E., Glynn P., Sigman K.* Queues with negative arrivals // J. Appl. Prob. — 1991. — V. 28. — P. 245–250.
2. *Bocharov P. P., D'Apice C., Manzo R., Pechinkin A. V.* Analysis of the Multi-Server Markov Queuing System with Unlimited Buffer and Negative Customers. // Automation and Remote Control. — 2007. — V. 68, No. 1. — P. 85–94.
3. *Kreinin A.* Queueing Systems with Renovation. // Journal of Applied Math. Stochast. Analysis. — 1997. — Vol. 10, No. 4. — P. 431–443.
4. *Kreinin A.* Inhomogeneous Random Walks: Applications in Queueing and Finance // CanQueue 2003. Fields Institute. Toronto.
5. *Bocharov P. P., Zaryadov I. S.* Queueing Systems with Renovation. Stationary Probability Distribution. // Bulletin of Peoples Friendship University of Russia. Series Mathematics. Information Sciences. Physics. — 2007. — No 1-2. — P. 14–23.
6. *Zaryadov I. S., Pechinkin A. V.* Stationary Time Characteristics of the GI/M/n/ System with Some Variants of the Generalized Renovation Discipline. // Automation and Remote Control. — 2009. — No. 12. — P. 2085–2097.
7. *Zaryadov I. S.* Stationary service characteristics of queueing system G/M/n/r with general renovation // Bulletin of Peoples Friendship University of Russia. Series Mathematics. Information Sciences. Physics. — 2008. — No. 2. — P. 3–10.
8. *Zaryadov I. S.* Queueing Systems with General Renovation. // International Congress on Ultra Modern Telecommunications, St. Petersburg: IEEE. — 2009. — P. 1–4.
9. *Zaryadov I. S.* The GI/M/n/ Queueing System with Generalized Renovation. // Automation and Remote Control. — 2010. — No. 4. — P. 663–671.
10. *Bocharov P.P., D'Apice C., Pechinkin A.V., Salerno S.* Queueing Theory. — VSP, Utrecht, Boston, 2004.

Author index

A

Abaev P. O. 9, 17
 Abrosimov L. I. 26, 34
 Adkhamova A. S. 41
 Aliev T. I. 45
 Andreev S. 301

B

Begishev V. O. 56
 Belyaeva J. O. 60
 Beschastny V. A. 9, 17
 Blinov A. I. 65
 Bogatyrev V. A. 73, 264
 Bogoiavlenskaia O. I. 281
 Bolotova G. O. 81
 Borodina A. B. 89
 Broner V. I. 349

D

Dao T. N. 92
 Daraseliya A. V. 408
 Demidova A. V. 150, 158
 Devyatkov V. V. 101
 Dinh T. D. 114
 Dolgushev R. A. 122
 Dudin M. S. 435

E

Eferina E. G. 129
 Efrosinin D. V. 89

F

Fedorov S. L. 463
 Fedorova E. 342
 Fomin M. B. 137

G

Gaidamaka Yu. V. 56
 Gerasimenko M. A. 143
 Gevorkyan M. N. 150, 158
 Gilmutdinov M. 481

Glukhov I. V. 165
 Gorbunova A. V. 169, 467
 Gorshenin A. K. 176
 Grebeshkov A. Yu. 184

H

Houankpo H. G. K. 473
 Hudoev I. 371
 Hussein O. A. 191

I

Ivanov A. E. 137
 Ivanov R. E. 435
 Izmaylova Y. 198

K

Kalimulina E. Yu. 203
 Kalinina K. A. 212
 Kanzitdinov S. K. 216
 Kaspirovich I. E. 225
 Kiporov D. S. 184
 Kirichek R. V. 114, 122, 388, 394,
 401
 Kocheganov V. M. 232
 Kokshenev V. 240
 Kolbasova V. 248
 Kolechkin A. O. 256
 Kolomoitcev V. S. 264
 Korolev V. Yu. 176
 Korolkova A. V. ... 129, 150, 158,
 427
 Korshok E. O. 272
 Korzun D. G. 281
 Koucheryavy A. 289
 Kovalchukov R. 301
 Kozyrev D. V. 473
 Kradenyh A. A. 169
 Kulik V. A. 305, 388
 Kulyabov D. S. 129, 150, 158, 427
 Kuznetsova O. V. 129

L

Larin A. A. 34
Larionov A. A. 435
Lisovskaya E. 248, 313

M

Makolkina M. A. ... 289, 326, 419
Matskevich I. A. 458
Matushenko S. I. 467
Melikov A. 452
Merzlyakov V. 381
Mikheev P. 240
Milovanova T. A. 445
Moiseev A. 452
Moiseeva S. 248, 313
Moltchanov D. A. 56, 301
Morozov E. V. 89, 334, 337
Morozov E. 337

N

Naung M. T. 101
Nazarov A. A. 342, 349, 356

O

Okuneva D. V. 191, 363
Orlov Y. N. 165

P

Pagano M. 313
Paramonov A. I. 92, 289
Parshutina S. A. 73
Peshkova I. 337
Petrov V. I. 56
Pham V. D. 388
Pirmagomedov R. 371
Pomortseva N. A. 356
Potakhina L. V. 334
Proshutinskiy K. S. 363

R

Razumchik R. V. 489
Rudenkova M. A. 26
Rumyantsev A. 337

S

Samouylov K. E. 9, 17, 301
Samuylov A. K. 56, 301
Scherbanskaya A. A. 458
Schetinin E. Yu. 381
Serebryakova A. A. 388
Sevastianov L. A. ... 65, 129, 150,
158, 427
Shangina D. 371
Shilin P. A. 394
Shklyaeva A. V. 401
Skubachevskii A. L. 41, 60
Sopin E. S. 408, 467
Sosnin V. V. 45, 412
Surodeeva E. V. 326
Suschenko S. 240

T

Teltevskaya V. A. 419
Tkachev R. 240
Tsarev A. S. 9

U

Udin D. N. 81
Ustyuzhanin N. 481

V

Vasilyev S. A. ... 65, 81, 216, 272
Vdovenko A. S. 281
Velieva T. R. 427
Vishnevsky V. M. 435
Vladyko A. G. 256
Vybornova A. I. 305

Y

Yapo G. S. 445
Yarkina N. V. 408

Z

Zadiranova L. 452
Zaryadov I. S. 169, 445, 458, 467,
489
Zorine A. V. 232
Zuev A. V. 184

Научное издание

**РАСПРЕДЕЛЕННЫЕ КОМПЬЮТЕРНЫЕ
И ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ:
УПРАВЛЕНИЕ, ВЫЧИСЛЕНИЕ, СВЯЗЬ
(DCCN-2016)**

В трех томах

**Том 3
Молодежная школа-семинар**

Издание подготовлено в авторской редакции

Технический редактор *Н.А. Ясько*
Дизайн обложки *М.В. Рогова*

Подписано в печать 07.11.2016 г. Формат 60×84/16.
Бумага офсетная. Печать офсетная. Гарнитура Таймс.
Усл. печ. л. 27,67. Тираж 500 экз. Заказ 1580

Российский университет дружбы народов
115419, ГСП-1, г. Москва, ул. Орджоникидзе, д. 3

Типография РУДН
115419, ГСП-1, г. Москва, ул. Орджоникидзе, д. 3, тел. 952-04-41

Для заметок

Для заметок
