View metadata, citation and similar papers at core.ac.uk
brought to you by CORE

МАССОВОЕ ОБСЛУЖИВАНИЕ:

ПОТОКИ, СИСТЕМЫ, СЕТИ

Материалы международной научной конференции
«Современные вероятностные методы анализа и оптимизации информационно-телекоммуникационных сетей»
Минск, 31 января – 3 февраля 2011 г.

QUEUES

FLOWS, SYSTEMS, NETWORKS

Proceedings of the International Conference
«Modern Probabilistic Methods for Analyisis and Optimization of Information and Telecommunication Networks»
Minsk, January 31 – February 3, 2011

Выпуск 21

Белорусский государственный университет

МАССОВОЕ ОБСЛУЖИВАНИЕ

ПОТОКИ, СИСТЕМЫ, СЕТИ

Материалы международной научной конференции «Современные вероятностные методы анализа и оптимизации информационно-телекоммуникационных сетей»
Минск, 31 января – 3 февраля 2011 г.

QUEUES

FLOWS, SYSTEMS, NETWORKS

Proceedings of the International Conference «Modern Probabilistic Methods for Analyisis and Optimization of Information and Telecommunication Networks»
Minsk, January 31 – February 3, 2011

МИНСК
«РИВШ»
2011

УДК 519.872:004.7(063)
ББК 32.973.202я43 + 22.18я43
М31

Рекомендовано кафедрой теории вероятностей и математической статистики БГУ
(протокол № 8 от 14 декабря 2010 г.)

Сборник основан в 1985 году

Редакционная коллегия:

д-р физ.-мат. наук, проф. А. Н. Дудин (отв. ред.);
д-р физ.-мат. наук В. И. Клименок (зам. отв. ред.);
канд. физ.-мат. наук Г. В. Царенков (отв. секр.);
д-р физ.-мат. наук, проф. Г. А. Медведев;
д-р физ.-мат. наук, проф. А. М. Андронов;
д-р физ.-мат. наук, проф. В. В. Анисимов;
д-р физ.-мат. наук, проф. В. В. Рыков;
д-р физ.-мат. наук, проф. Ю. В. Малинковский

М31 **Массовое** обслуживание : потоки, системы, сети : материалы международ. науч. конф. «Современные вероятностные методы анализа и оптимизации информационно-телекоммуникационных сетей», Минск, 31 янв. – 3 февр. 2011 г. Вып. 21 / редкол. : А. Н. Дудин (отв. ред.) [и др.]. – Минск : РИВШ, 2011. – 283 с.
ISBN 978-985-500-244-5

В сборнике излагаются новые результаты научных исследований в области разработки и оптимизации моделей процессов передачи информации в телекоммуникационных сетях с использованием аппарата теории систем и сетей массового обслуживания.

Предназначен для специалистов в области вероятностного анализа, случайных процессов, математического моделирования и математической статистики, а также для специалистов в области проектирования и эксплуатации сетей связи и компьютерных сетей.

ISBN 978-985-500-244-5

© БГУ, 2011
© Оформление. ГУО «Республиканский институт высшей школы», 2011

Настоящий сборник содержит материалы
ДВАДЦАТЬ ПЕРВОЙ БЕЛОРУССКОЙ ЗИМНЕЙ ШКОЛЫ-СЕМИНАРА ПО ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ (BWWQT-2011).
Минск, 31 января – 3 февраля 2011 г.

Школы-семинары организуются кафедрой теории вероятностей и математической статистики и научно-исследовательской лабораторией прикладного вероятностного анализа Белорусского государственного университета с 1985 года. Белорусские школы-семинары традиционно проводятся как научные конференции под различными девизами, отражающими основную идею проведения очередной конференции.
В 2011 году школа проводилась как международная конференция «Современные вероятностные методы анализа и оптимизации информационно-телекоммуникационных сетей».

ОРГАНИЗАЦИОННЫЙ КОМИТЕТ:

А. Дудин (Беларусь), председатель;
Г. Медведев (Беларусь), М. Ньютс (США), сопредседатели;
А. Андронов (Латвия), В. Анисимов (Великобритания),
Д. Баум (Германия), Х. Аль-Бегайн (Великобритания),
С. Чакраварти (США), Б. Чои (Корея),
Г. Фалин (Россия, Великобритания), А. Гомес-Корал (Испания),
Ч. Ким (Корея), В. Клименок (Беларусь), И. Коваленко (Украина),
А. Латков (Латвия), Е. Лебедев (Украина),
Ю. Малинковский (Беларусь), М. Маталыцкий (Беларусь),
А. Назаров (Россия), Р. Нобель (Нидерланды), В. Рыков (Россия),
С. Степанов (Россия), Я. Штрик (Венгрия), Х. Тиймс (Нидерланды),
В. Вишневский (Россия), А. Меликов (Азербайджан).

The present volume contains papers of
the TWENTY FIRST BELARUSIAN WINTER WORKSHOP
ON QUEUEING THEORY (BWWQT-2011).
Minsk, January 31 – February 3, 2011.

The workshops are organized by both department of the probability theory & mathematical statistics and research laboratory of the applied probability analysis of the Belarusian State University since 1985. The Belarusian workshops are traditionally conducted as scientific conferences under the various devices that reflect main idea of realization of the next conference.
In a 2011 the workshop was conducted as International conference «Modern Probabilistic Methods for Analyisis and Optimization of Information and Telecommunication Networks».

ORGANIZATIONAL COMMITTEE:

A. Dudin (Belarus), chairman;
G. Medvedev (Belarus), M. Neuts (USA), cochairmen;
A. Andronov (Latvia), V. Anisimov (United Kingdom),
D. Baum (Germany), K. Al-Begain (United Kingdom),
S. Chakravarthy (USA), B. Choi (Korea),
G. Falin (Russia, United Kingdom), A. Gomez-Corral (Spain),
C. Kim (Korea), V. Klimenok (Belarus), I. Kovalenko (Ukraine),
A. Latkov (Latvia), E. Lebedev (Ukraine), Y. Malinkovsky (Belarus),
A. Nazarov (Russia), R. Nobel (Netherlands), V. Rykov (Russia),
S. Stepanov (Russia), J. Sztrik (Hungury), H. Tijms (Netherlands),
V. Vishnevsky (Russia), A. Melikov (Azerbaijan).

Белорусские школы-семинары по теории массового обслуживания проводятся ежегодно, начиная с 1985 г. (раз в два года с 1999 г.) как научные конференции по различным приложениям теории массового обслуживания.

Предыдущими конференциями были:

1985	Гродно	Исследование путей повышения эффективности сетей связи и сетей ЭВМ
1986	Гомель	Математические методы теории массового обслуживания и их программное обеспечение
1987	Гродно	Принципы и алгоритмы коммутационных систем
1988	Гродно	Современные методы исследования информационно-вычислительных систем
1989	Гродно	Методы исследования информационно-вычислительных сетей
1990	Витебск	Математические методы исследования сетей связи и сетей ЭВМ
1991	Гродно	Сети связи и сети ЭВМ как модели массового обслуживания
1992	Брест	Сети связи и сети ЭВМ. Анализ и применение.
1993	Минск	Математические методы исследования систем и сетей массового обслуживания
1994	Минск	Анализ и применение систем и сетей массового обслуживания
1995	Минск	Исследование сетей связи и компьютерных сетей методами теории массового обслуживания
1996	Гродно	Исследование систем и сетей массового обслуживания
1997	Минск	Математические методы исследования телекоммуникационных сетей
1998	Минск	Математические методы исследования сетей массового обслуживания
1999	Минск	Современные математические методы исследования телекоммуникационных сетей
2001	Минск	Современные математические методы исследования информационно-вычислительных сетей
2003	Гомель	Современные математические методы анализа и оптимизации телекоммуникационных сетей
2005	Минск	Математические методы повышения эффективности функционирования телекоммуникационных сетей
2007	Гродно	Математические методы повышения эффективности информационно-телекоммуникационных сетей
2009	Минск	Современные математические методы анализа и оптимизации информационно-телекоммуникационных сетей

По составу участников проводимые школы-семинары до 1994 г. имели характер всесоюзных. Начиная с 1995 г. в работе школ-семинаров участвуют представители других стран (Австрии, Алжира, Бельгии, Болгарии, Венгрии, Германии, Нидерландов, Индии, Испании, Италии, Канады, Китая, Кореи, Мексики, Польши, Португалии, США, Турции, Франции, Японии).

The Belarusian workshops on queueing theory are conducted annually since 1985 (biennially since 1999) as scientific conferences on various applications of queueing theory.

The previous conferences were:

1985	Grodno	Research of paths of increase of efficiency of communication networks and computer networks
1986	Gomel	Mathematical methods of theory of queues and their software
1987	Grodno	Principles and algorithms of communication systems
1988	Grodno	Modern methods of a research of information systems
1989	Grodno	Methods of a research of information networks
1990	Vitebsk	Mathematical methods of a research of communication networks and computer networks
1991	Grodno	Communication networks and computer networks as models of a queuieng theory
1992	Brest	Communication networks and computer networks. The analysis and application.
1993	Minsk	Mathematical methods of a research of systems and queueing networks
1994	Minsk	Analysis and application of queueing systems and queueing networks
1995	Minsk	Research of communication networks and computer networks by methods of queueing theory
1996	Grodno	Research of queueing systems and queueing networks
1997	Minsk	Mathematical methods of a research of telecommunication networks
1998	Minsk	Mathematical methods of a research of queueing networks
1999	Minsk	Modern mathematical methods of investigation of telecommunication networks
2001	Minsk	Modern mathematical methods of investigation of the information networks
2003	Gomel	Modern mathematical methods of analysis and optimization of telecommunication networks
2005	Minsk	Mathematicals Methods of Optimization of Telecommunication Networks
2007	Grodno	Mathematical Methods for Increasing Efficiency of Information Telecommunication Networks
2009	Minsk	Mathematical Methods for Analysis and Optimization of Information Telecommunication Networks

On composition of participants the past workshops up to 1994 had the Soviet Union character. Since 1995 the representatives of other countries (Algeria, Austria, Belgium, Bulgaria, Canada, China, France, Germany, Hungary, India, Italy, Japan, Korea, Mexico, the Netherlands, Poland, Portugal, Spain, Turkey, the USA) participate in workshops.

ЛОКАЛЬНЫЙ ОРГКОМИТЕТ:

А. Н. Дудин (председатель), В. И. Клименок (зам. председателя), Г. А. Медведев, Ю. В. Малинковский, М. А. Маталыцкий, А. В. Бабицкий, В. В. Мушко, Г. В. Царенков, А. В. Казимирский, С. А. Дудин, О. С. Тарамин, Б. Сунь.

Материалы отпечатаны методом прямого репродуцирования оригинальных авторских текстов.

Для удобства читателей в конце помещен авторский указатель.

Адрес оргкомитета:
Беларусь, 220030, Минск, проспект Независимости, 4
Белорусский государственный университет
Факультет прикладной математики и информатики
Дудин Александр Николаевич, тел. +375 (0)17 2095486
Эл. почта: dudin@bsu.by
Интернет: http://www.bwwqt.bsu.by, http://www.apa.bsu.by

LOCAL ORGANIZING COMMITTEE:

A. Dudin (chairman), V. Klimenok (co-chairman), G. Medvedev, Yu. Malinkovski, M. Matalytsky, A. Babitski, V. Mushko, G. Tsarenkov, A. Kazimirsky, S. Dudin, O. Taramin, B. Sun.

The papers are printed by a method direct duplication of the original author's texts.

For convenience of the readers the author's index is located at the end.

Address of Organizing Committee:
Belarus, 220030, Minsk, 4 Pr. Nezavisimosty,
The Belarusian State University
Faculty of Applied Mathematics and Computer Science
Alexander Dudin, ph. +375 (0)17 2095486
E-mail: dudin@bsu.by
Internet: http://www.bwwqt.bsu.by, http://www.apa.bsu.by

ИССЛЕДОВАНИЕ СУММАРНОГО ПОТОКА ОБРАЩЕНИЙ В СИСТЕМЕ M|GI|∞ С ПОВТОРНЫМ ОБСЛУЖИВАНИЕМ С УЧЕТОМ НОМЕРА ПОПЫТКИ

И. Ананина

Томский государственный университет
Томск, Российская Федерация
ananinaia@sibmail.com

В работе рассматривается система M|GI|∞, на вход которой поступает простейший поток. В системе реализуется многократное обслуживание заявок, при этом длительность обслуживания и вероятность возвращения в систему зависят от номера обращения этой заявки. Найдена производящая функция суммарного числа заявок, обратившихся к системе за время t.

Ключевые слова: M|GI|∞ , потоки обращений, повторное обслуживание, метод предельной декомпозиции.

1. ПОСТАНОВКА ЗАДАЧИ

Рассматривается система массового обслуживания с неограниченным количеством обслуживающих приборов, на вход которой поступает простейший с параметром λ поток. В системе реализуется многократное обслуживание заявок. Заявка, выполняющая k-ое по счету обслуживание, называется k-заявкой. Первичные заявки, то есть заявки входящего простейшего потока, являются 1-заявкой. Завершив обслуживание, с вероятностью r_k покидает систему, а с вероятностью $1 - r_k$ возвращается на прибор для повторного обслуживания, становясь $(k+1)$-заявкой. Время обслуживания k-заявки имеет произвольную функцию распределения $B_k(x)$.

Обозначим $n_k(t)$ - число k-заявок, реализованных (появившихся) в системе за время t. Таким образом, имеем систему массового обслуживания с неограниченным числом обслуживающих приборов и повторными обращениями с учетом номера попытки . Ставится задача нахождения производящей функции суммарного процесса

$$m(t) = \sum_{k=1}^{\infty} n_k(t).$$

2. МЕТОД ПРЕДЕЛЬНОЙ ДЕКОМПОЗИЦИИ

Исследование суммарного потока обращений и нахождение его производящей функции будем осуществлять с помощью метода предельной декомпозиции [1]. Суть этого метода заключается в следующем.

Входящий поток делится по полиномиальной схеме с равными вероятностями на N независимых пуассоновских потоков с интенсивностями λ/N, заявки каждого потока направляются для обслуживания на соответствующий прибор. Таким образом, получаем совокупность N независимых однолинейных СМО. Будем полагать, что эти СМО с отказами. То есть новая заявка, поступившая в систему, занятую обслуживанием, теряется. При $N \to \infty$ вероятностью потерь заявок можно пренебречь, и тогда суммарные характеристики совокупности N однолинейных СМО сходятся к характеристикам исходной модели. Таким образом, задача нахождения распределения вероятностей числа обращений в СМО с неограниченным количеством обслуживающих приборов сводится к решению задачи нахождения распределения вероятностей числа обращений в однолинейной СМО с отказами.

3. ИССЛЕДОВАНИЕ ОДНОЛИНЕЙНОЙ СМО С ОТКАЗАМИ И ПОВТОРНЫМ ОБСЛУЖИВАНИЕМ С УЧЕТОМ НОМЕРА ПОПЫТКИ

Рассмотрим однолинейную систему обслуживания с отказами. В указанной однолинейной системе будем рассматривать соответствующий случайный процесс - суммарное число обращений, реализованных за время t:

$$m(t,N) = \sum_{k=1}^{\infty} n_k(t,N).$$

Здесь $n_k(t,N)$ - процессы изменения числа обращений k-заявок в однолинейной СМО. Поскольку рассматриваемый процесс $m(t,N)$ не является марковским, введем в рассмотрение дополнительные переменные:

процесс $k(t)$ - состояние обслуживающего прибора, то есть $k(t) = k$ когда он занят обслуживанием k-заявки , и $k(t) = 0$ когда прибор свободен;

$z(t)$ - длина интервала от момента времени t до момента окончания текущего обслуживания, если прибор занят.

Полученный трехмерный процесс $\{k(t), m(t,N), z(t)\}$ является марковским. Его распределение вероятностей обозначим следующим образом:

$P_0(m,t,N) = P\{k(t) = 0, m(t,N) = m\}$ - вероятность того, что в момент времени t прибор свободен и за это время к системе обратилось m заявок;

$P_k(m,z,t,N) = P\{k(t) = k, m(t,N) = m, z(t) < z\}$ - вероятность того, что на момент времени t в системе: суммарное число обращений равно m, на приборе осуществляется обслуживание , и до его окончания остается времени меньше z. Для распределения вероятностей трехмерного марковского процесса $\{k(t), m(t,N), z(t)\}$ составим прямую систему дифференциальных уравнений Колмогорова [2]:

$$\frac{\partial P_0(m,t,N)}{\partial t} = \sum_{k=1}^{\infty}(1-r_k)\frac{\partial P_k(m,z,t,N)}{\partial z} - \frac{\lambda}{N}P_0(m,t,N),$$

$$\frac{\partial P_1(m,z,t,N)}{\partial t} = \frac{\partial P_1(m,z,t,N)}{\partial z} - \frac{\partial P_1(m,0,t,N)}{\partial z} + \frac{\lambda}{N}P_0(m-1,t,N)B_1(z),$$

$$\frac{\partial P_k(m,z,t,N)}{\partial t} = \frac{\partial P_k(m,z,t,N)}{\partial z} - \frac{\partial P_k(m,0,t,N)}{\partial z} + r_{k-1}B_k(z)\frac{\partial P_{k-1}(m-1,0,t,N)}{\partial z}.$$

Рассмотрим производящие функции

$$\sum_{m=0}^{\infty} x^m P_0(m,t,N) = H_0(x,t,N), \tag{1}$$

$$\sum_{m=0}^{\infty} x^m P_k(m,z,t,N) = H_k(x,z,t,N),\ k \geq 1. \tag{2}$$

Перейдем к системе дифференциальных уравнений в частных производных [3] для функций $H_0(x,t,N)$, $H_k(x,z,t,N)$, $k \geq 1$ вида

$$\frac{\partial H_0(x,t,N)}{\partial t} = \sum_{k=1}^{\infty}(1-r_k)\frac{\partial H_k(x,z,t,N)}{\partial z} - \frac{\lambda}{N}H_0(x,t,N),$$

$$\frac{\partial H_1(x,z,t,N)}{\partial t} = \frac{\partial H_1(x,z,t,N)}{\partial z} - \frac{\partial H_1(x,0,t,N)}{\partial z} + \frac{\lambda}{N}H_0(x,t,N)xB_1(z), \tag{3}$$

$$\frac{\partial H_k(x,z,t,N)}{\partial t} = \frac{\partial H_k(x,z,t,N)}{\partial z} - \frac{\partial H_k(x,0,t,N)}{\partial z} + r_{k-1}xB_k(z)\frac{\partial H_{k-1}(x,0,t,N)}{\partial z}.$$

Решение системы (3) будем искать в виде

$$H_0(x,t,N) = 1 + \frac{1}{N}F_0(x,t) + o(N^{-2}), \tag{4}$$

$$H_k(x,z,t,N) = \frac{1}{N}F_k(x,z,t) + o(N^{-2}),\ k \geq 1. \tag{5}$$

Тогда уравнения для $F_0(x,t)$ и $F_k(x,z,t)$, $k \geq 1$, имеют вид:

$$\frac{\partial F_0(x,t)}{\partial t} = -\lambda + \sum_{k=1}^{\infty}(1-r_k)f_k(x,t), \tag{6}$$

$$\frac{\partial F_1(x,z,t)}{\partial t} = \frac{\partial F_1(x,z,t)}{\partial z} - f_1(x,t) + \lambda x B_1(z), \tag{7}$$

$$\frac{\partial F_k(x,z,t)}{\partial t} = \frac{\partial F_k(x,z,t)}{\partial z} - f_k(x,t) + r_{k-1}xB_k(z)f_{k-1}(x,t), \tag{8}$$

где

$$f_k(x,t) = \frac{\partial F_k(x,0,t)}{\partial z},\ k \geq 1.$$

Решение дифференциального уравнения первого порядка в частных производных (7) определяется решением системы обыкновенных дифференциальных уравнений для характеристических кривых

$$\frac{dt}{1} = \frac{dz}{-1} = \frac{dF_1(x,z,t)}{\lambda x B_1(z) - f_1(x,t)}.$$

Найдем два первых интеграла этой системы. Один из них найдем из уравнения

$$-dt = dz.$$

То есть

$$z = C_1 - t. \tag{9}$$

Другой первый интеграл находим из уравнения

$$dF_1(x,z,t) = [\lambda x B_1(z) - f_1(x,t)]dt.$$

Откуда, подставляя (9), получаем равенство

$$F_1(x,z,t) = \Phi(z+t) + \int_0^t [\lambda x B_1(z+t-s) - f_1(x,s)]ds, \tag{10}$$

где $\Phi(z)$ - произвольная дифференцируемая функция.

Для определения частного решения необходимо воспользоваться начальными условиями

$$P_0(m,0,N) = \begin{cases} 0, & m > 0, \\ R_0(N), & m = 0. \end{cases}$$

$$P_k(m,z,0,N) = \begin{cases} 0, & m > 0, \\ R_k(z,N), & m = 0. \end{cases} \tag{11}$$

$$R_0(N) + \sum_{k=1}^{\infty} R_k(z,N) = 1.$$

Здесь $R_0(N)$, $R_k(z,N)$, $k \geq 1$, - стационарное распределение вероятностей состояния прибора в однолинейной СМО. Из (1-2) следует, что выполняются равенства

$$H_0(1,t,N) = R_0(N),\ H_k(1,z,t,N) = R_k(z,N), \tag{12}$$

Стационарные вероятности $R_0(N)$, $R_k(z,N)$ можно представить в виде

$$R_0(N) = 1 + \frac{1}{N}R_0 + o(N^{-2}),\ R_k(z,N) = \frac{1}{N}R_k(z) + o(N^{-2}). \tag{13}$$

Тогда согласно (4-5), (10) и (11)

$$F_1(x,z,0) = \Phi(z) = R_1(z).$$

Вероятности R_0, $R_k(z)$, $k \geq 1$, с учетом (12) определяются решением системы (6-8) при $x = 1$ и условием нормировки, то есть

$$R_0 = -\lambda\left(b_1 + \sum_{k=2}^{\infty}\prod_{i=1}^{k-1} r_i b_{i+1}\right),$$

$$R_1(z) = \lambda\int_0^z (1 - B_1(y))dy, \tag{14}$$

$$R_k(z) = \lambda\prod_{i=1}^{k-1} r_i \int_0^z (1 - B_k(y))dy,$$

где b_k - математическое ожидание времени обслуживания k-заявки. Таким образом, частное решение уравнения (7) принимает вид

$$F_1(x, z, t) = \lambda\int_0^{z+t} (1 - B_1(y))dy + \int_0^t [\lambda x B_1(z + t - s) - f_1(x, s)]ds. \tag{15}$$

Дифференцируя это тождество по z в нуле, получаем интегральное уравнение для нахождения неизвестной функции $f_1(x, t)$:

$$f_1(x, t) = \lambda(1 - B_1(t)) + \lambda x B_1(t). \tag{16}$$

Таким же образом найдем решение уравнений (8):

$$F_k(x, z, t) = \lambda\prod_{i=1}^{k-1} r_i \int_0^{z+t} (1 - B_k(y))dy +$$

$$+ \int_0^t [r_{k-1} x B_k(z + t - s) f_{k-1}(x, s) - f_k(x, s)]ds. \tag{17}$$

Дифференцируя полученное тождество (17) для $F_k(x, z, t)$, $k \geq 2$, по z в нуле, получаем интегральное уравнение для нахождения функций $f_k(x, t)$, $k \geq 2$:

$$f_k(x, t) = \lambda\prod_{i=1}^{k-1} r_i(1 - B_k(t)) + x r_{k-1}\int_0^t b_k(t - s) f_{k-1}(x, s)ds, \tag{18}$$

где $b_k(t)$ - плотность распределения времени обслуживания k-заявки.

Соответственно решение уравнения (6) имеет вид

$$F_0(x, t) = -\lambda\left(b_1 + \sum_{k=2}^{\infty}\prod_{i=1}^{k-1} r_i b_{i+1}\right) - \lambda t + \sum_{k=1}^{\infty}(1 - r_k)\int_0^t f_k(x, s)ds. \tag{19}$$

При $z \to \infty\infty$ выражения (15) и (17) приобретают вид

$$F_1(x, t) = \lambda b_1 + \lambda x t - \int_0^t f_1(x, s)ds, \tag{20}$$

$$F_k(x,t) = \lambda b_k \prod_{i=1}^{k-1} r_i + r_{k-1} x \int_0^t f_{k-1}(x,s)ds - \int_0^t f_k(x,s)ds. \qquad (21)$$

Тогда, суммируя (19-21) по маргинальной компоненте k, согласно (4-5), получаем

$$F(x,t) = -F_0(x,t) + \sum_{k=1}^{\infty} F_k(x,t) = (x-1)\left[\lambda t + \sum_{k=1}^{\infty} r_k \int_0^t f_k(x,s)ds\right]. \qquad (22)$$

Производящая функция $G(x,t)$ числа суммарных обращений в исходной системе с неограниченным числом линий определяется выражением

$$G(x,t) = \mathbf{e}^{\{} Mm(t)\} = \lim_{N\to\infty} \mathbf{e}^{\{} M(m_1(t) + m_2(t) + \ldots + m_N(t))\}.$$

В силу того, что все m_i, $i = 1 \ldots N$, независимы и одинаково распределены, имеем

$$G(x,t) = \lim_{N\to\infty} (H(x,t,N))^N,$$

где $H(x,t,N)$ - производящая функция суммарного числа обращений в однолинейной СМО. Учитывая (4-5) получаем

$$G(x,t) = \lim_{N\to\infty} \left(1 + \frac{1}{N} F(x,t) + o(N^{-2})\right)^N.$$

Подставляя выражение (22), получаем искомое выражение, определяющее производящую функцию числа суммарных обращений в системе M|GI|∞ с повторным обслуживанием с учетом номера попытки:

$$G(x,t) = \mathbf{e}$$

$$\left\{(\text{x-1})\left[\lambda t + \sum_{k=1}^{\infty} r_k \int_0^t f_k(x,s)ds\right]\right\}.$$

Знание вида производящей функции суммарного числа обращений в системе позволяет расчитать числовые характеристики процесса $m(t)$ и определить, например, среднее число обращений к системе за время наблюдения t.

ЛИТЕРАТУРА

1. *Морозова А.С., Моисеева С.П., Назаров А.А.* Исследование СМО с повторным обращением и неограниченным числом обслуживающих приборов методом предельной декомпозиции // Вычислительные технологии.– 2005.– Т.13, вып.5.– С. 88-92.
2. *Гнеденко Б.В.* Курс теории вероятностей. М.: Наука, 1969.
3. *Эльцгольц Л.Э.* Дифференциальные уравнения и вариационное исчисление. М.: Наука, 1969.

AN UNRECURRENT FLOW WITH TRIANGULAR CONDITIONAL DISTRIBUTION OF INTERARRIVAL TIMES

A. Andronov

Transport and Telecommunication Institute
Riga, Latvia
lora@mailbox.riga.lv

An unrecurrent flow is considered when interarrival times form a Markov chain. A conditional distribution of an interarrival time is a triangular one, for that a parameter is a linear function on a previous interarrival time. The stationary one-dimensional and two-dimensional distributions have been derived. It is shown that possible values of a correlation coefficient between the adjacent interarrival times are in the diapason (-1/3, 1/3).

Keywords: Unrecurrent flow, interarrival time, stationary distribution, dependence.

1. INTRODUCTION

Classical models of queueing, reliability, inventory and etc., suppose Poisson or recurrent flows of considered input claims. But numerous statistical data prove the opposite. For example, it has been experimentally stated that interarrival times of Internet flows are the dependent ones [5]. Analogously, flows of insurance claims for damages have dependent structure [4]. In the first case, a dependence between interarrivals is described by the so-called Batch Markovian Arrival Process [7, 8, 9, 10, 11, 13] or Markov-Additive Processes of Arrivals [12]. In the second case, copulas are used usually for a description of the dependence [1, 2, 3, 4, 11].

In our paper we would like to use natural and direct way for a description of the dependence between continue random variables $X_0, X_1, ...$ those are interpreted as interarrival times of flow claims. We suppose that the last ones correspond to a Markov chain with continue state space $\Omega = (0, 1)$. To describe a transit probability for one step we consider a family $\Psi = \{f_\theta(x) : x, \theta \in \Omega\}$ of triangular distributions that is determined by a probability density function

$$f_\theta(x) = \begin{cases} 2\frac{x}{\theta}, & 0 < x < \theta, \\ 2\frac{1-x}{1-\theta}, & \theta \le x < 1. \end{cases} \tag{1}$$

The corresponding cumulative distribution function is calculated as

$$F_\theta(x) = \begin{cases} \frac{1}{\theta}x^2, & 0 < x < \theta, \\ 1 - \frac{(1-x)^2}{1-\theta}, & \theta \le x < 1. \end{cases} \tag{2}$$

Now for Markov chain $X_1, X_2, ..., X_n, ...$ we suppose the following: if $X_n = x$, then X_{n+1} has a distribution from family Ψ with parameter $\theta = \theta(x)$. In the previous paper of the author [2] a case where $\theta = x$ has been considered. It turns out, that a correlation coefficient between X_n and X_{n+1} equals 1/3. A case $\theta = 1 - x$ gives a value -1/3.
Now we consider a case $\theta = \delta(1-x) + x(1-\delta) = \delta + x - 2\delta x$ where δ is a parameter, $0 < \delta < 1$. Below we show that all value between - 1/3 and 1/3 can be received in such a way.
Our paper has the following structure. At first we find marginal stationary cumulative $G(x) = P\{X \le x\}$ and density $g(x) = \frac{\partial}{\partial x}G(x)$ distribution function for this Markov chain; then two-dimensional stationary density function $g(x,y) = \frac{\partial^2}{\partial x \partial y}P\{X_n \le x, X_{n+1} \le y\}$ and its mixed moment. Numerical illustrations and conclusion end the paper.

2. STATIONARY MARGINAL DISTRIBUTION

We would like to present calculating probability density function $g(x) = G(x)'$ for the stationary marginal distribution. Further we consider a case $\delta < 0.5$. Let us consider two consecutive terms of stationary Markov chain. If $X_n = z$ then X_{n+1} has a distribution from family Ψ with parameter $\theta = \delta(1-z) + z(1-\delta) = \delta + z - 2\delta z$. We see that always $\theta \in (\delta, 1-\delta)$. An inequality $X_{n+1} = x < \theta = \delta + z - 2\delta z$ implies $z > (x-\delta)/(1-2\delta)$. Therefore from (1) we get for $\delta < x < 1-\delta$

$$g(x) = \int\limits_{(x-\delta)/(1-2\delta)}^{1} g(z)\frac{2x}{\delta + z - 2\delta z}dz + \int\limits_{0}^{(x-\delta)/(1-2\delta)} g(z)2\frac{1-x}{1-(\delta+z-2\delta z)} =$$

$$= 2x\int\limits_{(x-\delta)/(1-2\delta)}^{1} g(z)\frac{1}{\delta + z - 2\delta z}dz + 2(1-x)\int\limits_{0}^{(x-\delta)/(1-2\delta)} g(z)\frac{1}{1-(\delta+z-2\delta z)}dz. \quad (3)$$

Further

$$\frac{\partial}{\partial x}g(x) = 2\int\limits_{(x-\delta)/(1-2\delta)}^{1} g(z)\frac{1}{\delta + z - 2z\delta}dz - 2xg\Big(\frac{x-\delta}{1-2\delta}\Big)\frac{1}{1-2\delta}\frac{1}{x} -$$

$$-2\int\limits_{0}^{(x-\delta)/(1-2\delta)} g(z)\frac{1}{1-(\delta+z-2z\delta)}dz + 2(1-x)g\Big(\frac{x-\delta}{1-2\delta}\Big)\frac{1}{1-2\delta}\frac{1}{1-x},$$

$$\frac{\partial^2}{\partial x^2}g(x) = -2g\Big(\frac{x-\delta}{1-2\delta}\Big)\frac{1}{1-2\delta}\frac{1}{x} - 2\frac{1}{1-2\delta}\frac{\partial}{\partial x}g\Big(\frac{x-\delta}{1-2\delta}\Big) -$$

$$-\frac{2}{1-x}g\Big(\frac{x-\delta}{1-2\delta}\Big)\frac{1}{1-2\delta} + 2\frac{1}{1-2\delta}\frac{\partial}{\partial x}g\Big(\frac{x-\delta}{1-2\delta}\Big). \quad (4)$$

Finally for $\delta < x < 1-\delta$

$$x(1-x)\frac{\partial^2}{\partial x^2}g(x) = -2g\Big(\frac{x-\delta}{1-2\delta}\Big)\frac{1}{1-2\delta}. \tag{5}$$

For $0 < x < \delta$ we have from (3)

$$g(x) = 2x\int_0^1 g(z)\frac{1}{\delta+z-2\delta z}dz. \tag{6}$$

If we denote

$$a = 2\int_0^1 g(z)\frac{1}{\delta+z-2\delta z}dz \tag{7}$$

then

$$g(x) = ax, \quad 0 < x < \delta. \tag{8}$$

For $x > 1-\delta$ we have analogously

$$g(x) = 2(1-x)\int_0^1 g(z)\frac{1}{1-(\delta+z-2\delta z)}dz. \tag{9}$$

Obviously $g(x)$ is a symmetric function with respect to $1/2$: $g(x) = g(1-x)$, $1/2 < x < 1$. Then

$$2\int_0^1 g(z)\frac{1}{1-(\delta+z-2\delta z)}dz = 2\int_0^1 g(1-z)\frac{1}{\delta+(1-z)-2\delta(1-z)}dz = a,$$

$$g(x) = a(1-x), \quad 1-\delta \le x \le 1. \tag{10}$$

3. CALCULATING PROCEDURE

The equation (4) is not convenient for a calculation. It is better to use equation (3). When can be calcalated the last integral in (3) for fixed x using $g(z)$ for $z < x$? Obviously when $(x-\delta)/(1-2\delta) < x$, therefore for $x < 1/2$. To have the same possibility for the other integral, we apply equality (6): if $\delta < x$, then

$$\int_{(x-\delta)/(1-2\delta)}^{1} g(z)\frac{1}{\delta+z(1-2\delta)}dz = \frac{a}{2} - \int_0^{(x-\delta)/(1-2\delta)} g(z)\frac{1}{\delta+z(1-2\delta)}dz,$$

$$g(x) = ax - 2x\int_0^{(x-\delta)/(1-2\delta)} \frac{g(z)}{\delta+z(1-2\delta)}dz + 2(1-x)\int_0^{(x-\delta)/(1-2\delta)} \frac{g(z)}{1-\delta-z(1-2\delta)}dz. \tag{11}$$

We can use this formula for consequence calculations. Above, we see that on interval $(0,\delta)$ density $g(x)$ has a known view. Let us consider number sequence

$$b_0=\delta,\ b_i=\delta+b_{i-1}(1-2\delta)=\delta\sum_{j=0}^{i}(1-2\delta)^j=\frac{1}{2}\big(1-(1-2\delta)^{i+1}\big),\quad i=1,2,\dots\ . \quad (12)$$

Equation (10) shows that values of $g(x)$ on interval (b_i,b_{i+1}) were calculated via $g(x)$'s previous values on $(0,b_i)$. A border value $1/2$ is arrived for $i=\infty$, so we must end calculations when b_i is close to $1/2$.
One should note that a numerical calculation by formula (10) for $x<1/2$ is very simple. During the recurrent calculation we set $a=1$. On leaving one, constant a is determined from normalization condition for $g(x)$: a probabilistic mass on interval $(0,1/2)$ equals $1/2$. For $1/2<x<1$ we use equalities $g(x)=g(1-x)$ and $G(x)=1-G(1-x)$.
Let us perform analytical calculations for $x\in(b_0,b_1)=(\delta,2\delta(1-\delta))$. If $z<\delta$ then $g(z)=az$, so

$$g(x)=ax-2x\int\limits_0^{(x-\delta)/(1-2\delta)}\frac{az}{\delta+z(1-2\delta)}dz+2(1-x)\int\limits_0^{(x-\delta)/(1-2\delta)}\frac{az}{1-\delta-z(1-2\delta)}dz=$$
$$=ax-\frac{2ax}{1-2\delta}\Big[\frac{x-\delta}{1-2\delta}-\frac{\delta}{1-2\delta}ln\Big(\frac{x}{\delta}\Big)\Big]-\frac{2a(1-x)}{1-2\delta}\Big[\frac{x-\delta}{1-2\delta}+\frac{1-\delta}{1-2\delta}ln\Big(\frac{1-x}{1-\delta}\Big)\Big]=$$
$$=ax-2a\frac{x-\delta}{(1-2\delta)^2}+2ax\frac{\delta}{(1-2\delta)^2}ln\Big(\frac{x}{\delta}\Big)-2a(1-x)\frac{1-\delta}{(1-2\delta)^2}ln\Big(\frac{1-x}{1-\delta}\Big)$$

Therefore for $x\in(b_0,b_1)=(\delta,2\delta(1-\delta))$

$$g(x)=ax-\frac{2a}{(1-2\delta)^2}\Big\{x-\delta-x\delta ln\Big(\frac{x}{\delta}\Big)+(1-x)(1-\delta)ln\Big(\frac{1-x}{1-\delta}\Big)\Big\}. \quad (13)$$

Corresponding cumulative distribution function for $x\in(b_0,b_1)=(\delta,2\delta(1-\delta))$

$$G(x)=\frac{1}{2}a\Big(\frac{x-\delta}{1-2\delta}\Big)^2+\frac{1}{2}ax^2+(1-2x)\frac{a}{1-2\delta}\frac{x-\delta}{1-2\delta}+$$
$$+x^2\frac{a\delta}{(1-2\delta)^2}ln\Big(\frac{x}{\delta}\Big)+(1-x)^2\frac{a(1-\delta)}{(1-2\delta)^2}ln\Big(\frac{1-x}{1-\delta}\Big)\Big\}. \quad (14)$$

An extension of two last formulas on interval $x\in(b_0,0.5)=(\delta,0.5)$ will be named *the first-order approximation* $g_1(x)$ and $G_1(x)$ correspondingly. As $G_1(0.5)=0.5$ we determine normalizing constant a from (13) :

$$a^{-1}=\frac{1}{2}-\frac{1}{2}\frac{\delta}{(1-2\delta)^2}\ln(\delta)-\frac{1-\delta}{2(1-2\delta)^2}\ln(1-\delta)-\frac{1}{2(1-2\delta)^2}\ln(2). \quad (15)$$

Our experience shows that for $\delta>0.1$ the first-order approximation ensures a sufficient precision of the calculation. The subsequent approximations are determined

analogously. As border value 1/2 from (11) is arrived for $i = \infty$, we must always restrict ourselves by some approximation.
Moments $\mu_r = E(X^r)$ are calculated by a usual way throw density function $g(x)$. Obviously $\mu_1 = 0.5$. The other moments $\{\mu_r\}$ for the first-order approximation $g_1(x)$ are given in Appendix.

4. STATIONARY TWO-DIMENSIONAL DISTRIBUTION

Stationary probability density function $g(x,y) = \frac{\partial^2}{\partial x \partial y} P\{X \le x, Y \le y\}$ is expressed via the stationary density $g(x)$ and transition probability density function (1). We must consider case $\theta = x + \delta - 2x\delta > y$ and contrary case separately:

$$g(x,y) = g(x) f_{\delta + x(1-2\delta)}(y) = \begin{cases} g(x)\frac{2y}{\delta + x(1-2\delta)}, & y \le \delta + x(1-2\delta), \\ g(x)\frac{2(1-y)}{1-\delta-x(1-2\delta)}, & otherwise. \end{cases} \tag{16}$$

As it is shown in the Appendix, he second mixed moment is calculated as

$$E(XY) = \frac{1}{6}\Big[1 + \delta + 2(1-2\delta)\mu_2\Big]. \tag{17}$$

Now correlation coefficient $\rho = (E(XY) - 0.25)/D(X)$ can be calculated.

5. NUMERICAL RESULTS

In this section we give some numerical comparing. Table 1 shows how different approximations of the density function behave oneself for small value of parameter δ. Three functions are presented here for $\delta = 0.01$: the first-order approximation $g1(x)$, the second-order approximation $g2(x)$, and a curve of function $g0(x)$, that gives precise values for $\delta = 0$ (see [3]). We see, there exists an appreciable difference between these real values and the values received by the first-order approximation. For $\delta = 0.3$ the first-order and the second-order approximations fully coincide, therefore they are precise.

Table 1 **Comparing of different approximations for small parameter δ**

x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
$q0(x)$	0.285	0.540	0.765	0.960	1.125	1.260	1.365	1.440	1.485
$q2(x)$	0.277	0.530	0.757	0.954	1.122	1.260	1.369	1.488	1.499
$q1(x)$	0.292	0.560	0.797	1.000	1.166	1.293	1.377	1.414	1.402

The Table 2 contains values of stationary density functions for a sum of two adjacent interarrivals for $\delta = 0.005$. Row $fSum(x)$ corresponds to density (15). Row $fInd(x)$ corresponds to independent interarrival times with density $g2(x, 0.01)$. One can see that positive correlation is appreciable.
The correlation coefficient is a linear function of $\delta : \rho(\delta) = -\frac{2}{3}(\delta - 0.5)$, $0 \le \delta \le 1$. So,

for $0 \le \delta \le 1/2$ we have a positive correlation dependence, for $1/2 \le \delta \le 1$ - a negative one.
We see that the considered dependence has some appreciable influence on indices of interest.

Table 2 **Comparing of density functions for a sum**

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$fSum(x)$	0.028	0.108	0.226	0.367	0.515	0.655	0.781	0.882	0.950	0.975
$fInd(x)$	0.006	0.043	0.129	0.264	0.435	0.619	0.808	0.986	1.122	1.178

6. CONCLUSION

We have considered an approach for appreciating dependence between interarrival times. It is supposed that interarrival times correspond to Markov chain with triangular distribution of transition probabilities. Expressions for stationary one-dimensional and two-dimensional distributions of interarrival times and corresponding numerical characteristics have been derived. The following investigations will be devoted to a modification of the used form of a triangular distribution to expand a possible diapason of correlation coefficient values.

7. APPENDIX

Firstly, we derive the expressions for moments $\mu_r = E(X^r)$, $r = 2, 3, ...$, for the first-order approximation $g_1(x)$. Let us introduce the following notations:

$$\Delta(k,c,d) = \int_c^d x^k ln(x)dx = \frac{1}{k+1}d^{k+1}ln(d) - \frac{1}{k+1}c^{k+1}ln(c) - \frac{1}{k+1}\Big(d^{k+1} - c^{k+1}\Big),$$

$$v_r = \int_0^{1/2} x^r g_1(x)dx.$$

Then

$$v_r = \int_0^{1/2} x^r g_1(x)dx = \int_0^{1/2} x^r axdx + \frac{2a}{(1-2\delta)^2}\int_\delta^{1/2}[\delta + (1-\delta)ln(1-\delta)]x^r dx -$$

$$- \frac{2a}{(1-2\delta)^2}\int_\delta^{1/2}[1 + \delta ln(\delta) + (1-\delta)ln(1-\delta)]x^{r+1}dx +$$

$$+ \frac{2a\delta}{(1-2\delta)^2}\int_\delta^{1/2} x_{r+1}ln(x)dx - \frac{2a(1-\delta)}{(1-2\delta)}\int_\delta^{1/2}(1-x)ln(1-x)x^r dx =$$

$$= a\frac{1}{r+2}\Big(\frac{1}{2}\Big)^{r+2} + \frac{2a}{(1-2\delta)^2}[\delta + (1-\delta)ln(1-\delta)]\frac{1}{r+1}\Big(\Big(\frac{1}{2}\Big)^{r+1} - \delta^{r+1}\Big) -$$

$$-\frac{2a}{(1-2\delta)^2}[1+\delta ln(\delta)+(1-\delta)ln(1-\delta)]\frac{1}{r+2}\Big(\Big(\frac{1}{2}\Big)^{r+2}-\delta^{r+2}\Big)+$$

$$+\frac{2a\delta}{(1-2\delta)^2}\Delta\Big(r+1,\delta,\frac{1}{2}\Big)+\frac{2a(1-\delta)}{(1-2\delta)^2}\sum_{i=0}^{r}\binom{r}{i}(-1)^{r-i}\Delta\Big(r-i+1,1-\delta,\frac{1}{2}\Big).$$

Further

$$\mu_r=\int_0^1 x^r g_1(x)dx=\int_0^{1/2} x^r g_1(x)dx+\int_{1/2}^1 x^r g_1(1-x)dx=$$

$$=\int_0^{1/2} x^r g_1(x)dx-\int_{1/2}^0 (1-y)^r g_1(y)dy=\int_0^{1/2} x^r g_1(x)dx+\sum_{i=0}^{r}(-1)^i\binom{r}{i}\int_0^{1/2} y^i g_1(y)dy=$$

$$=v_r+\sum_{i=0}^{r}(-1)^i\binom{r}{i}v_i.$$

$$\mu_1=v_0=\frac{1}{2},\quad \mu_2=2v_2-2v_1+\frac{1}{2}.$$

Secondly, we derive formula (16) for a mixed moment:

$$E(XY)=\int_0^1\int_0^{\delta+x(1-2\delta)}\frac{g(x)2y}{\delta+x(1-2\delta)}xydydx+\int_0^1\int_{\delta+x(1-2\delta)}^1\frac{g(x)2(1-y)}{1-[\delta+x(1-2\delta)]}xydydx=$$

$$=\frac{2}{3}\int_0^1 g(x)(\delta+x(1-2\delta))^2xdx+$$

$$+\int_0^1\frac{2x}{1-[\delta+x(1-2\delta)]}g(x)\frac{1}{6}(3y^2-2y^3)dx\Big|_{y=\delta+x(1-2\delta)}^{y=1}=$$

$$=\frac{2}{3}\int_0^1 g(x)(\delta+x(1-2\delta))^2xdx+\frac{1}{3}\int_0^1\frac{x}{1-[\delta+x(1-2\delta)]}g(x)\Big\{1-(\delta+x(1-2\delta))^2-$$

$$-2[\delta+x(1-2\delta)]^2[1-[\delta+x(1-2\delta)]]\Big\}dx=\frac{2}{3}\int_0^1 g(x)(\delta+x(1-2\delta))^2xdx+$$

$$+\frac{1}{3}\int_0^1 x[1+\delta+x(1-2\delta)]g(x)dx-\frac{2}{3}\int_0^1 xg(x)[\delta+x(1-2\delta)]^2dx=$$

$$=\frac{1}{3}[(1+\delta)\mu_1+(1-2\delta)\mu_2]=\frac{1}{3}\Big[(1+\delta)\frac{1}{2}+(1-2\delta)\mu_2\Big].$$

REFERENCES

1. *Abegaz F., Naik-Nimbalkar U. V.* Modelling statistical dependence of Markov chains via copula models // Journal of statistical planning and inference. 2008. № 138(4). P. 1131–1146.
2. *Andronov A.* A triangular symmetric copula // The Third International Conference Problems of Cybernetics and Informatics . 2010. V. II. P. 36–40.
3. *Andronov A* On a copula for failure times of system elements // Mathematical and statistical models and methods in reliability. V.V Rykov et al. (eds.), Springer Science + Business Media, LLC. 2011. P. 39–51.
4. *Embrechts P. , Lindskog F., McNeil A.* Modelling Dependence with Copulas and Applications to Risk Management (Chapter 8) // Handbook of Heavy Tailed Distributions in Finance, Amsterdam, Elsevier. 2003. P. 329–384.
5. *Hernandez-Campos F., Jeffay K. F., Park C., Marron J. S., Resnick S. I.* Extremal dependence: Internet traffic applications // Stochastic Models. 2005. V. 22. № 1. P. 1–35.
6. *KHeyman D. P., Lucantoni D.* Modeling multiple IP traffic streams with rate limits. // IEE/ACM Transactions on Networking (TON). 2003. V. 11. № 6. P. 948–958.
7. *Kim C. S., Klimenok V., Taramin O.* A tandem retrial queueing system with two Markovian flows and reservation of channels // Computers and Operations Research. 2010. V. 37. № 7. P. 1238–1246.
8. *Kim C. S., Klimenok V., Mushko V., Dudin A.* The BMAP/PH/N retrial queueing system operating in Markovian random environment // Computers and Operations Research. 2010. V. 37. № 7. P. 1228–1237.
9. *Kim C. S., Park S. H., Dudin A., Klimenok V., Tsarenkov G.* Investigaton of the BMAP/G/1,/PH/1/M tandem queue with retrials and losses // Applied Mathematical Modelling. 2010. V. 34. № 10. P. 2929–2940.
10. *Kim C. S., Dudin A., Klimenok V., Khramova V.* Erlang loss queueing system with batch arrivals operating in a random environment // Computers and Operations Research. 2009. V. 36. № 3. P. 674–697.
11. *Nelsen B. R.* An Introduction to Copulas. Second Edition New York: Springer. 2006.
12. *Pacheco A., Tang L. C., Prabhu U. N.* Markov-Modulated processes and Semiregenerative Phenomena New Jersey - London - Singapore: World Scientific. 2009.
13. *Scott S. L., Smyth P.* The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling // Bayesian Statistics. Oxford University Press. 2003. V. 7. P. 1–10.

VERIFICATION OF STABILITY REGION OF A RETRIAL QUEUING SYSTEM BY REGENERATIVE METHOD

K. Avrachenkov[1], R. Goricheva[2], E. Morozov[2]

[1] *INRIA,*

[2] *IAMR KarSC RAS*

[1]*Sophia-Antipolis, France,*

[2]*Petrozavodsk, Russia*

Konstantin.Avratchenkov@sophia.inria.fr, lana–@mail.ru, emorozov@karelia.ru

Regenerative processes play an important role in the modeling and simulation of the modern telecommunication systems. In this note, we apply the regenerative approach to estimate the steady-state blocking probability in a single-class retrial system with constant retrial rate. Moreover, this estimation allows to verify stability region of this system.

This work is supported by Russian Foundation for Basic Research under grant 10-07-00017.

Keywords: retrial queueing system; regenerative simulation; blocking probability; confident estimation; stability condition

1. DESCRIPTION OF THE MODEL AND ITS STABILITY CONDITION

We consider a $M/G/1/1$-type retrial system with Poisson input with rate λ and general service time with rate μ ($\mathsf{E}S = 1/\mu$). Arrivals who find the server busy join the infinite capacity orbit, and then return to the system after exponentially distributed retrial time with rate μ_0. Thus, the total input stream to the server consists of two (generally, dependent) streams: a Poisson $\lambda-$input of primary customers and the input of retrial (or orbit customers) with a rate $\tilde{\mu}_0 \leq \mu_0$, where $\tilde{\mu}_0 = \mu_0$ when the orbit is not empty, and $\tilde{\mu}_0 = 0$, otherwise, because this input may have gaps when the orbit is empty. We denote this original system as Σ.

This retrial model is motivated by modelling telephone exchange systems [5], ALOHA type multiple access protocols [4], and short TCP transfers [2]. In particular, the model with one server and no buffering considered in the present work represents well an ALOHA type multiple access channel. If there are n retrying customers, each customer retries with the rate μ_0/n.

Note that, following [1], analysis below can be extended to a wider class of retrial systems containing more that one server, non-zero buffer for waiting customers, and a general renewal $\lambda-$ input of primary customers.

We also construct an auxiliary single-server loss (with no buffer) system $\hat{\Sigma}$ as follows. The system $\hat{\Sigma}$ has the same Poisson $\lambda-$input, the service time but, unlike original system, it also has an independent Poisson input of customers with rate μ_0 (we call them μ_0-customers). Moreover, arriving customer in the system $\hat{\Sigma}$ who finds the server busy *leaves the system forever and do not affect the future state.*

It is easy to see that the server in the original system is less loaded since the actual input rate from the orbit to the server $\tilde{\mu}_0 \leq \mu_0$. It is expected that the distribution of the states (busy or empty) of system Σ approaches the corresponding distribution of system $\hat{\Sigma}$ provided the orbit size in Σ increases. It shows that the stability condition of the orbit which is fed by the rejected customers in system $\hat{\Sigma}$ must guarantee stability of the orbit in (less loaded) system Σ. Indeed this fact has been proved in [1].

Specifically, denote by P_{loss} the stationary loss probability in system $\hat{\Sigma}$ (which always exists under our assumptions). Then it has been proved in [1] that the following condition

$$(\lambda + \mu_0)\mathsf{P}_{loss} < \mu_0, \tag{1}$$

is sufficient stability condition of the orbit in system Σ.

In particular, it implies finiteness of the mean regeneration period of the system (to be defined below) in continuous and discrete time.

There are some systems for which explicit expressions for the probability P_{loss} is known. In particular, if system $\tilde{\Sigma}$ is the c-server Erlang model (that is M/G/c/c loss system), then stability condition (1) takes the form

$$\frac{\Big((\lambda+\mu_0)/\mu\Big)^c}{c!}\left[\sum_{n=0}^{c}\frac{\Big((\lambda+\mu_0)/\mu\Big)^n}{n!}\right]^{-1} < \frac{\mu_0}{\lambda+\mu_0}. \tag{2}$$

It allows us to write the stability condition (1) for $M/G/1/1$ system under consideration in an explicit form as follows:

$$\frac{\lambda+\mu_0}{\lambda+\mu_0+\mu} < \frac{\mu_0}{\lambda+\mu_0}. \tag{3}$$

2. REGENERATIVE ANALYSIS OF THE RETRIAL SYSTEM

In this section, we consider the regenerative approach to simulate original retrial system and, in particular, describe various regenerative structures of the original system. Denote by $N(t)$ the number of retrial customers (orbit size) at instant t and let $\nu(t)$ be the state of server (0 or 1) at instant t. Let $\{t_n\}$ be arrival instants of λ-customers in both systems, and denote $N(t_n^-) = N_n$, $\nu(t_n^-) = \nu_n$, $n \geq 1$. Also denote $X = \{X(t) =$

$N(t) + \nu(t)$, $t \geq 0\}$, and let $X(t_n^-) = X_n := N_n + \nu_n$, $n \geq 1$. We note that system $\hat{\Sigma}$ regenerates at the instants when the λ-customers find the server empty, because at these instants we may use the memoryless property of the input of μ_0-customers. (This procedure can be extended to a general renewal λ-input, see [1].) Thus, regenerations of the discrete-time basic process $\{X_n\}$ are defined as follows. Let $\beta_0 = 0$, then

$$\beta_{n+1} = \inf_k (k > \beta_n : X_k = 0),\ n \geq 0, \tag{4}$$

are required (classical) regenerations. (Construction of regenerations for continuous-time process X is evident.) Denote by $R(t)$ the total number of rejected customers in system Σ in interval $[0, t]$, and let $A(t)$ be the total number of calls (primary λ-customers and μ_0-customers) in interval $[0, t]$. Denote also by R and A the number of orbit customers and total number of arrivals, respectively, during the regeneration cycle. It has been proved in [1] that under condition (3) the process $\{R(t),\ t \geq 0\}$ is positive recurrent with the embedded regenerations $\{\beta_n\}$ and, in particular, there exists the limit with probability 1 (w.p.1)

$$\lim_{t\to\infty} \frac{R(t)}{A(t)} = \frac{\mathsf{E}R}{\mathsf{E}A}. \tag{5}$$

Let $I_n = 1$ if the nth customer is rejected (otherwise, $I_n = 0$). It is easy to show that the regeneration period is aperiodic and hence, the weak limit $I_n \Rightarrow I$ exists. Moreover, (by uniform integrability of indicators) the following convergence holds

$$\mathsf{P}(I_n = 1) \to \mathsf{P}_{orb}, \qquad n \to \infty, \tag{6}$$

where $\mathsf{P}_{orb} := \mathsf{E}I$ is the stationary probability to join the orbit. Also by the standard result of regenerative theory,

$$\hat{\mathsf{P}}_{orb}(n) := \frac{\sum_{k=1}^n I_k}{n} \to \frac{\mathsf{E}R}{\mathsf{E}A} = \mathsf{P}_{orb}, \quad n \to \infty, \tag{7}$$

that is the sample mean estimator of probability P_{orb} converges w. p. 1 to P_{orb}. Thus, both limit ratios (see also (5)) are consistent and give the same expression for stationary probability P_{orb}.

Because original system is less loaded than $\hat{\Sigma}$, then in the stability region one can expect that the basic process X is positive recurrent (that is $\mathsf{E}\beta < \infty$) and thus the following convergence and inequality hold:

$$\hat{\mathsf{P}}_{orb}(n) \to \mathsf{P}_{orb} \leq \mathsf{P}_{loss}, \quad \text{as} \quad n \to \infty. \tag{8}$$

On the other hand, in the instability region, an unlimited increase of the orbit size is expected. Under this condition, the output of retrial customers (going from orbit to server) approaches to the Poisson input with rate μ_0 that is $\tilde{\mu}_0 \to \mu_0$, and we expect that the estimator $\hat{\mathsf{P}}_{orb}(n)$ must approach P_{loss}. But in this case the process X *is not positive recurrent regenerative*, and the existence of the limit in (8) is, in general, an

open problem. Hopefully, we can use another type of regenerations, more exactly, *quasi-regenerations*, to estimate P_{orb} in this case. As quasi-regenerations of original (unstable) system we can take the instants when arriving primary customer meets an empty server (while the orbit size *may be arbitrary*). Namely, take $\alpha_0 = 0$, then quasi-regenerations are define as follows:

$$\alpha_{n+1} = \inf_k (k > \alpha_n : \nu_k = 0),\ n \geq 0. \tag{9}$$

It follows from the previous discussion that "more" the system is unstable (that is if parameters λ, μ, μ_0 are taken "deeper" in the instability region), the less difference $\mathsf{P}_{loss} - \hat{\mathsf{P}}_{orb}(n)$ must be in the limit as $n \to \infty$. In this regard, it is important to note that in the instability region classical regenerations are expected to be rare and to terminate finally. More correctly, we expect that in the instability region, after a finite (w.p.1) time t_0, the orbit will never be empty, and thus after this instant the original system completely couples with the auxiliary system $\tilde{\Sigma}$ fed by the two independent $\lambda-$ and μ_0-inputs. In other words, quasi-regenerations become classical regenerations but for the process $\{\nu_n\}$ considered in isolation. As a result, the estimate $\hat{\mathsf{P}}_{orb}(n)$ will converge to P_{loss} w.p.1 as $n \to \infty$. A difference between quasi-regenerations and classical ones are expected to appear at the stability boundary and around the boundary.

The purpose of this note is to check by regenerative (or quasi-regenerative) simulation stability/instability of the original system, in particular, comparing estimate $\hat{\mathsf{P}}_{orb}(n)$ with explicit formula for P_{loss}.

Note that detailed description of regenerative method can be found in [3, 6, 7].

3. SIMULATION RESULTS

In this section, we present some simulation results related to the estimation of the blocking probability in the original system both in stability and instability regions. We consider the $M/G/1/1$ system with the service time to be either exponential or Pareto and assuming $\lambda = 1$. Recall that our main goal is to verify stability of the original system by simulation assuming that stability condition (3) holds. (Recall that condition (3) relates formally to system $\hat{\Sigma}$.)

Note that for the service time with Pareto distribution $\mathsf{P}(S > x) = (x/x_0)^{-\alpha}$, $x \geq x_0$ ($\mathsf{P}(S > x) = 1$, $x \leq x_0$), assumption $\alpha > 2$ implies finiteness of the 2nd moment, $\mathsf{E}S^2 < \infty$, and, as a result, the finiteness of the 2nd moment of the regeneration period, $\mathsf{E}\beta^2 < \infty$ (see [9]). Thus, the confidence estimation of the probability P_{orb} based on the regenerative simulation can be applied [6], although we present below only point estimation.

Denote by $\delta = \mathsf{P}_{loss} - \hat{\mathsf{P}}_{orb}(n)$ the difference between the estimator value and the explicit value of loss probability (see (2))

$$\mathsf{P}_{loss} = \frac{\lambda + \mu_0}{\lambda + \mu_0 + \mu}.$$

Denote also by Γ the difference between two sides of stability condition (3), that is

$$\Gamma = \frac{\mu_0}{\lambda + \mu_0} - \frac{\lambda + \mu_0}{\lambda + \mu_0 + \mu}.$$

Note that $\Gamma > 0$ in the stability region, $\Gamma < 0$ in the instability region, and $\Gamma = 0$ corresponds to the boundary of stability region. One expect that the value of the "distance" Γ closely related with the obtained estimate $\hat{\mathsf{P}}_{orb}(n)$ and hence with δ. Namely, we expect that the bigger $\Gamma > 0$ (in the stability region), the bigger is $\delta > 0$, while if $\Gamma \downarrow 0$, one can expect $\delta \downarrow 0$. On the other hand, we expect the approximation $\delta \approx 0$ in the instability region, where $\Gamma < 0$.

Simulation results for M/M/1/1 systems are presented in Table 1, where n is indeed a number of *regenerations* (4) in the stability region, where $\Gamma > 0$, and is the number of quasi-regenerations (9) in the instability region, where $\Gamma < 0$, obtained during the simulation.

Table 1: Simulation of M/M/1/1

μ_0	μ	n	P_{loss}	$\hat{\mathsf{P}}_{orb}(n)$	δ	Γ
2	4	10829	0,4286	0,302	0,1266	0,2381
4	3	8301	0,625	0,4685	0,1565	0,1750
6	2,5	6023	0,7368	0,5992	0,1376	0,1203
0,6	3	2059	0,3478	0,3325	0,0153	0,0272
2	1,6	946	0,6522	0,6388	0,0134	0,0145
6	1,3	705	0,8434	0,8275	0,0159	0,0138
2	1,4	6349	0,6818	0,6793	0,0025	-0,0152
4	1,1	3655	0,8197	0,8167	0,0030	-0,0197
0,5	2,3	11988	0,3947	0,3946	0,0001	-0,0614
4	0,1	385	0,9804	0,9804	0,0000	-0,1804
0,1	2	12910	0,3548	0,3537	0,0011	-0,2639
0,1	0,1	1748	0,9167	0,9125	0,0042	-0,8258

Similar results are obtained for M/Pareto/1/1 systems with Pareto service time distribution $\mathsf{P}(S \geq x) = (x/x_0)^{-2,5}$, $x \geq x_0$, see Table 2. Note that we put the rows in the tables in the correspondence with the "distance" from the stability boundary, expressed by the value of Γ.

Note that stability can also be empirically verified by observation that there is no tendency in the dynamics of the orbit size, while instability effect is expressed by visual unlimited growth of the orbit size. Indeed, the dynamics of the orbit size is illustrated by Figure 1 when stability condition (3) holds, while Figure 2 shows this dynamics when this condition is violated.

Table 2: Simulation of M/Pareto/1/1

μ_0	x_0	n	P_{loss}	$\hat{\mathsf{P}}_{orb}(n)$	δ	Γ
2,7	0,1	14335	0,9569	0,2057	0,7512	0,3483
1	0,1	12715	0,9231	0,1808	0,7423	0,2500
2	0,2	7809	0,9000	0,4033	0,4967	0,1667
2,2	0,4	710	0,8276	0,6631	0,1645	0,0066
0,6	0,22	423	0,8136	0,3666	0,4470	0,0053
1,65	0,37	339	0,8112	0,5986	0,2126	0,0023
1	0,36	9022	0,7692	0,5470	0,2222	-0,0455
2	0,5	5652	0,7826	0,7159	0,0667	-0,0476
0,1	0,4	11537	0,6226	0,4224	0,2002	-0,3322

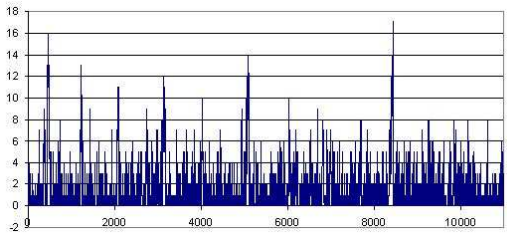

Fig. 1. Stable dynamics of orbit in M/M/1/1, $\mu_0 = 4$, $\mu = 3$.

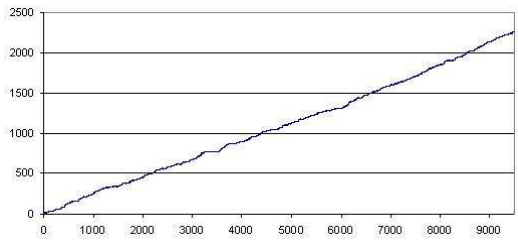

Fig. 2. Instability of orbit in M/M/1/1, $\mu_0 = 0,5$, $\mu = 2,3$.

Thus, the simulation results confirm the theoretical analysis. Moreover, in fact, we confirm by simulation that the given sufficient stability condition is also necessary one.

4. CONCLUSION

In this note, we consider a new retrial systems which can be used to describe, for instance, behavior of short TCP transfers or ALOHA type multiple access protocol. For a single-server system with no buffer and Poisson input we estimate blocking probability both in the stability region and in the instability region. In particular, in the stability region we use classical regenerative simulation, while in the instability region

we rely on the quasi-regenerations appearing when primary customers meet the empty server. Simulation of M/M/1/1 and M/Pareto/1/1 retrial systems show a remarkable consistence with the know analytical results for the Erlang type retrial systems.

REFERENCES

1. *Avrachenkov K., Morozov E. V.* Stability analysis of GI/G/c/K Retrial Queue with Constant Retrial Rate, INRIA, Research Report No 7335, July 2010 (Sophia Antipolis).
2. K. Avrachenkov and U. Yechiali, "Retrial networks with finite buffers and their application to Internet data traffic", *Probability in the Engineering and Informational Sciences*, v.22, pp.519-536, 2008.
3. *Bratley P. , Fox B. L.* A guide to simulation // New York: Springer-Verlag. 1987.
4. *Choi B.D., Rhee K.H., Park K.K.* The M/G/1 retrial queue with retrial rate control policy. // Probability in the Engineering and Informational Sciences. 1993. V. 7, P. 29-46.
5. *Fayolle G.* A simple telephone exchange with delayed feedback, In O.J. Boxma, J.W. Cohen and H.C. Tijms (eds.) Teletraffic Analysis and Computer Performance Evaluation. 1986. V. 7. P. 245-253,
6. *Glynn P. W., Iglehart D. L.* Conditions for the applicability of the regenerative method // Management Science. 1993. V. 39. P. 1108–1111.
7. *Law A. M. , Kelton W. D.* Simulation modeling and analysis // New York: McGraw- Hill . 1991.
8. *Morozov E. V.* Weak regeneration in modeling of queueing processes // Queueing Systems. 2004. V. 46. P. 295–315.
9. *Wolff R. W.* Stochastic Modeling and the Theory of Queues // Prentice-Hall . 1989.

OPTIMIZATION OF EMERGENCY INVENTORY REPLENISHMENT OF A PERISHABLE PRODUCT

A. Babitski

Belarusian State University
Minsk, Belarus
babitski@hotmail.com

Profitability of an additional replenishment cycle in an inventory system with a perishable product with deterministic shelf life, Poisson demand process, and fixed replenishment lead time is studied. Formulae for calculating income and expenses in the system are obtained. Behaviour of the net income as a function of ordering moment and order quantity is analyzed and an approach for determining parameter values that maximize net income is proposed.

Keywords: Inventory system, ordering moment optimization, order quantity optimization, stochastic demand, deterministic shelf life.

1. INTRODUCTION

Classical inventory systems, e.g., studied in [5] offer flexibility in determining optimal duration of the replenishment cycle based on particular values of storage and replenishment costs. In contrast to them the systems with perishable products have to deal with some time constraints stipulated by the assumptions about product deterioration, as in [6, 7, 8, 9].

Stochastic demand creates uncertainty about inventory level at some moment in the future, and a number of papers focus on optimization of system performance, binding ordering decisions to the moments when inventory reaches some level, see [1, 3, 4].

This paper studies a system with fixed-life product which can be sold up to some predefined time. Time intervals between depletion moments are assumed to be exponentially distributed. Incurred costs include item price, delivery and storage costs. Revenue is generated by sales. Unsold items can be utilized at some profit or some loss. Stochastic nature of demand implies that initial inventory may be insufficient to last till the end of the sales cycle, therefore a possibility of placement of an additional order with a fixed lead time is considered. Ordering moment and quantity are the parameters to be determined in the task for maximization of the net income in the system.

The proposed model is useful in describing midday production cycle of a highly perishable product in a grocery store or replenishment of a seasonal product or fashion apparel in the middle of season. In many situations precise estimation of demand during the future sale cycle is not available, what translates to the high probability to discard a lot of unsold products. A practical alternative is to start with some inventory

level corresponding to the pessimistic estimation of demand and to place a additional order later, depending on the actual intensity of sales. Particular benefit of determining the optimal order placement moment from the standpoint of grocery store production practice is that it facilitates in finding the optimal production schedule in the store.

2. MODEL

It is assumed that inventory level in the system at moment 0 is r. Items are consumed in accordance with Poisson process with intensity μ, generating income p per each item sold before moment T, after which all unsold items are utilized with income q. Holding cost is h per each item per unit time. There is a possibility to place an order for delivery of some l additional items with purchase cost d per each item and c cost for delivery. The order is delivered b time units after the ordering moment x. The purpose is to determine the optimal values of x and l that maximize total income in the system.

Storage cost during time t when starting inventory level was j can be calculated as

$$s_1(t) = \frac{h\left(1 - e^{-\mu x}\right)}{\mu},$$

$$s_j(t) = hjte^{-\mu t} + \int_0^t \left(hj\tau + s_{j-1}(t-\tau)\right) d\left(1 - e^{-\mu\tau}\right), j > 1.$$

Using Laplace transform, as in [2], it can be expressed explicitly as

$$s_j(t) = \frac{h}{\mu} \sum_{k=1}^{j} k \left(1 - \sum_{n=0}^{j-k} \frac{(\mu t)^n}{n!} e^{-\mu t}\right).$$

Applying this formula for intervals $[0, x+b]$ and $[x+b, T]$ the expression for total storage cost in the system can be written as

$$S_l(x) = \frac{h}{\mu} \sum_{k=1}^{r} k \left(1 - \sum_{n=0}^{r-k} \frac{(\mu(x+b))^n}{n!} e^{-\mu(x+b)}\right) +$$

$$+\frac{h}{\mu} \sum_{m=0}^{r} \psi_m^{(r)}(x+b) \sum_{k=1}^{l+m} k \left(1 - \sum_{n=0}^{l+m-k} \frac{(\mu(T-x-b))^n}{n!} e^{-\mu(T-x-b)}\right)$$

where $\psi_i^{(j)}(t)$ are the probabilities to have i items in the system after time t if the starting inventory was j, calculated as

$$\psi_i^{(j)}(t) = \begin{cases} \frac{(\mu t)^{j-i}}{(j-i)!} e^{-\mu t}, i > 0 \\ 1 - \sum_{n=0}^{j-1} \frac{(\mu t)^n}{n!} e^{-\mu t}, i = 0 \end{cases}$$

Replenishment expenses are $c + ld$.

The revenue is formed by the income from sales:

$$p\sum_{m=0}^{r}\psi_m^{(r)}(x+b)\sum_{i=0}^{m+l}(m+l-i)\psi_i^{(m+l)}(T-x-b)+p\sum_{m=0}^{r}(r-m)\psi_m^{(r)}(x+b)$$

and utilization income:

$$q\sum_{m=0}^{r}\psi_m^{(r)}(x+b)\sum_{i=0}^{m+l}i\psi_i^{(m+l)}(T-x-b)$$

Explicitly total revenue can be expressed as:

$$R_l(x)=p(r+l)-(p-q)e^{-\mu T}\sum_{n=0}^{r-1}\frac{(\mu(x+b))^n}{n!}\sum_{j=0}^{r-n+l-1}\frac{(\mu(T-x-b))^j}{j!}(r-n+l-j)-$$

$$-(p-q)\left(1-\sum_{n=0}^{r-1}\frac{(\mu(x+b))^n}{n!}e^{-\mu(x+b)}\right)\sum_{j=0}^{l-1}\frac{(\mu(T-x-b))^j}{j!}e^{-\mu(T-x-b)}(l-j)$$

3. OPTIMIZATION OF NET INCOME

Net income in the system accrued on time interval [0,T] can be expressed as:

$$I_l(x)=R_l(x)-S_l(x)-c-dl$$

Search for the optimal parameters l and x that maximize $I_l(x)$ can be limited to the area $l<=\mu(T-x-b)$ because net income in the system with $l>\mu(T-x-b)$ is less then net income in the system with parameter $l=\mu(T-x-b)$.

For each fixed value of l there exist single value of x that maximizes $I_l(x)$ because of its convexity in the aforementioned area, as $I_l''(x)<0$:

$$I_l''(x)=-(p-q)\mu^2e^{-\mu T}\left(\frac{(\mu(x+b))^{r-1}}{(r-1)!}\sum_{j=0}^{l-2}\frac{(\mu(T-x-b))^j}{j!}+\right.$$

$$\left.+\frac{(\mu(T-x-b))^{(l-1)}}{(l-1)!}\left(e^{\mu(x+b)}-\sum_{n=0}^{r-2}\frac{(\mu(x+b))^n}{n!}\right)\right)-$$

$$-\mu h\left[le^{-\mu(x+b)}\frac{(\mu(x+b))^{r-1}}{(r-1)!}-e^{-\mu T}\frac{(\mu(x+b))^{r-1}}{(r-1)!}\sum_{k=1}^{l-1}k\frac{(\mu(T-x-b))^{l-k-1}}{(l-k-1)!}+\right.$$

$$\left.+e^{-\mu T}\sum_{n=0}^{r-2}\frac{(\mu(x+b))^m}{m!}\sum_{k=1}^{l}\frac{(\mu(T-x-b))^{l-k}}{(l-k)!}-e^{-\mu(T-x-b)}\sum_{k=1}^{l}\frac{(\mu(T-x-b))^{l-k}}{(l-k)!}\right]$$

So that, optimal value of x can be found by iterations with the help of $I_l'(x)$:

$$I_l'(x)=(q-p)\mu\left[e^{-\mu(T-x-b)}\sum_{j=0}^{l-1}\frac{(\mu(T-x-b))^j}{j!}+\right.$$

$$+ e^{-\mu T} \sum_{j=0}^{l-1} \frac{(\mu(T-x-b))^j}{j!} \sum_{n=0}^{r-2} \frac{(\mu(x+b))^n}{n!} \Bigg] -$$

$$-h \Bigg[e^{-\mu T} \sum_{m=0}^{r-1} \frac{(\mu(x+b))^m}{m!} \sum_{k=1}^{l} k \frac{(\mu(T-x-b))^{l-k}}{(l-k)!} -$$

$$-e^{-\mu(T-x-b)} \sum_{k=1}^{l} k \frac{(\mu(T-x-b))^{l-k}}{(l-k)!} - e^{-\mu(x+b)} \sum_{m=0}^{r-1} \frac{(\mu(x+b))^m}{m!} \Bigg]$$

4. NUMERICAL EXAMPLE

Impact of various values of ordering moment and quantity on the net income in the inventory system with parameters $\mu = 2$, $d = 2$, $p = 9.5$, $b = 2$, $r = 13$, $q = 0.5$, $c = 3$, $h = 1.5$, $T = 12$ is shown in table 1.

Table 1 **Impact of ordering moment and quantity on the net income**

	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$	$l = 7$	$l = 8$
$x = 0.0$	52.09	51.20	49.44	46.78	43.16	38.52	32.79	25.93
$x = 0.5$	52.84	52.69	51.69	49.78	46.91	43.01	38.04	31.93
$x = 1.0$	53.59	54.19	53.93	52.77	50.64	47.49	43.26	37.90
$x = 1.5$	54.33	55.66	56.14	55.72	54.33	51.92	48.42	43.80
$x = 2.0$	55.04	57.10	58.29	58.59	57.92	56.22	53.44	49.53
$x = 2.5$	55.73	58.46	60.34	61.31	61.32	60.31	58.21	54.97
$x = 3.0$	56.35	59.71	62.22	63.82	64.45	64.06	62.58	59.95
$x = 3.5$	56.91	60.82	63.88	66.04	67.23	67.38	66.43	64.32
$x = 4.0$	57.38	61.77	65.31	67.93	69.58	70.18	69.65	67.91
$x = 4.5$	57.77	62.55	66.47	69.48	71.49	72.41	72.15	70.61
$x = 5.0$	58.08	63.16	67.38	70.67	72.92	74.02	73.83	72.27
$x = 5.5$	58.31	63.63	68.06	71.52	73.87	74.94	74.60	72.77
$x = 6.0$	58.48	63.96	68.52	72.02	74.27	75.08	74.32	71.98
$x = 6.5$	58.60	64.17	68.74	72.11	74.01	74.27	72.84	69.85
$x = 7.0$	58.67	64.25	68.68	71.65	72.89	72.32	70.07	66.46
$x = 7.5$	58.70	64.16	68.19	70.40	70.65	69.09	66.08	62.06
$x = 8.0$	58.64	63.75	66.96	68.02	67.08	64.60	61.13	57.10
$x = 8.5$	58.42	62.71	64.57	64.17	62.16	59.19	55.74	52.10
$x = 9.0$	57.74	60.43	60.47	58.80	56.28	53.43	50.48	47.49
$x = 9.5$	55.85	56.04	54.52	52.44	50.23	47.98	45.73	43.48
$x = 10.0$	50.68	49.18	47.68	46.18	44.68	43.18	41.68	40.18

Maximal value of net income 75.129 is achieved for ordering moment 5.8315 and ordering quantity 6.

REFERENCES

1. *Axsater S.* Evaluation of installation stock based (R,Q)-policies for two-level inventory systems with Poisson demand // Operations Research. 1998. V. 46. P. 135–145.
2. *Babitski A., Lee M.H.* Inventory level optimization in the conditions of fluctuating demand // 6th International Workshop on MSPT Proceedings. 2006. P. 15–16.
3. *Chen F.* Echelon reorder points, installation reorder points, and the value of centralized demand information // Management Science. 1998. V. 44. P. 221–234.
4. *Forsberg R.* Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand // European Journal of Operational Research. 1997. V. 96. P. 130–138.
5. *Hadley G., Whitin T.M.* Analysis of Inventory Systems // Prentice-Hall, Englewood Cliffs, NJ. 1963.
6. *Kanchanasuntorn K., Techanitisawad A.* An approximate periodic model for fixed-life perishable products in a two-echelon inventory-distribution system // International Journal of Production Economics. 2006. V. 100(1). P. 101–115.
7. *Nahmias S.* Perishable inventory theory // Operations Research. 1982. V. 30. P. 680–708.
8. *Pal M.* An inventory model for deteriorating items when demand is random // Calcutta Statistical Association Bulletin. 1990. V. 39. P. 201–207.
9. *Ravichandran N.* Probabilistic analysis of a continuous review perishable system with Markovian demand. Erlangian live and non-instanteneous lead times // OR Spektrum. 1988. V. 10. P. 23–27.

AN OPEN QUEUEING NETWORK WITH PARTLY NON-ACTIVE CUSTOMERS.

J. Bojarovich*, **Yu. Malinkovsky**

Gomel State University named after Skorina
Gomel, Belarus
* juls1982@list.ru

An open queueing network with partly non-active customers is considered. Non-active customers are in a system queue and do not get service. Customers may turn from non-active condition into condition, when they may get their service and vice versa. The form of stationary distribution and criterion of stationary distribution existence are obtained.

Keywords: partly non-active customers, reversibility, ergodicity, traffic equations, stationary distribution.

1. INTRODUCTION

Nowadays queueing networks with partly non-active customers become actual to a marked degree. Non-active customers are in a system queue and do not get service. Results for queueing networks with partly non-active customers may be used for epidemic progress analyzing and prognosis.

In paper [1] G. Tsitsiashvili and M. Osipova observed an open queueing network with non-active customers and established the form of stationary distribution.

This paper generalizes results for network from [1]. We consider network, where customers may partly loose their capacity for service. Customers may turn from non-active condition into condition, when they may get their service and vice versa. We researched the form of stationary distribution and established the criterion of stationary distribution existence.

2. SYSTEM DESCRIPTION

Consider isolated system. Positive and negative customers arrive at the system according to a Poisson processes at rates λ and λ^- accordingly. When arriving at system positive customer queues up and needs service. A negative customer induces a positive customer, if any, to leave the system immediately. A negative customer does not produce any action if there are no positive customers at the system. There is input Poisson flow of non-active customers at rate θ. Non-active customers queues up and can not get service. There are three input Poisson flows of signals at rates ν, ψ and φ. When arriving at the system the signal at rate ν induces an ordinary customer, if

any, to become non-active. When arriving at the system the signal at rate φ induces a non-active customer, if any, to become an ordinary. A signal at rate ψ induces a non-active customer, if any, to leave the system immediately. Signals do not need service. The service time has exponential distribution with the parameter μ.

Let $n(t)$ and $n'(t)$ are numbers of ordinary and non-active customers at the system at time t accordingly. $X(t) = (n(t), n'(t))$ is a continuous-time Marcov chain with a space of states $X = \{(n, n'),\ n, n' = 0, 1, \ldots\}$.

Under the condition of stationary distribution existence stationary probabilities $\pi(n, n')$ satisfy global balance equations:

1) $n = 0,\ n' = 0$:

$$(\lambda + \theta)\pi(n, n') = (\mu + \lambda^-)\pi(n + 1, n') + \psi\pi(n, n' + 1)\ ;$$

2) $n > 0,\ n' = 0$:

$$(\lambda + \mu + \lambda^- + \nu + \theta)\pi(n, n') = \lambda\pi(n - 1, n') + (\mu + \lambda^-)\pi(n + 1, n') +$$

$$+\varphi\pi(n - 1, n' + 1) + \psi\pi(n, n' + 1)\ ;$$

3) $n = 0,\ n' > 0$:

$$(\lambda + \varphi + \theta + \psi)\pi(n, n') = (\mu + \lambda^-)\pi(n + 1, n') + \nu\pi(n + 1, n' - 1) +$$

$$+\theta\pi(n, n' - 1) + \psi\pi(n, n' + 1)\ ;$$

4) $n > 0,\ n' > 0$:

$$(\lambda + \mu + \lambda^- + \nu + \varphi + \theta + \psi)\pi(n, n') = \lambda\pi(n - 1, n') + (\mu + \lambda^-)\pi(n + 1, n') +$$

$$+\nu\pi(n + 1, n' - 1) + \varphi\pi(n - 1, n' + 1) + \theta\pi(n, n' - 1) + \psi\pi(n, n' + 1)\ ;$$

Lemma. *Condition of reversibility is:*

$$\lambda\nu\psi = \theta(\mu + \lambda^-)\varphi.$$

Theorem 1. *If inequalities*

$$\lambda < \mu + \lambda^-,\ \ \theta < \psi,$$

hold, hence Marcov process $X(t)$ is ergodic. And under the condition of reversibility $X(t)$ has stationary distribution:

$$\pi(n, n') = \left(1 - \frac{\lambda}{\mu + \lambda^-}\right)\left(1 - \frac{\theta}{\psi}\right)\left(\frac{\lambda}{\mu + \lambda^-}\right)^n\left(\frac{\theta}{\psi}\right)^{n'}.$$

3. NETWORK DESCRIPTION

Consider open queueing network with set of systems $J = \{1, 2, \ldots, N\}$. Positive and negative customers arrive at the network according to a Poisson processes at rates λ_i and λ_i^- accordingly, $i \in J$. Non-active customers arrive at the network according to a Poisson processes at rates θ_i, $i \in J$. Non-active customers queue up to corresponding system and can not get service. There are input Poisson flows of signals at rates ν_i, φ_i и ψ_i, $i \in J$. When arriving at the system $i \in J$ the signal at rate ν_i induces an ordinary customer at system, if any, to become non-active. When arriving at the system $i \in J$ the signal at rate φ_i induces an non-active customer, if any, to become an ordinary. A signal at rate ψ_i induces an non-active customer, if any, to leave the system $i \in J$ immediately. Signals do not need service. Service times are independent exponentially distributed random values with parameters μ_i, $i \in J$.

After finishing of service process at system $i \in J$ customer is routed to system $j \in J$ with the probability $p_{i,j}$ as a positive customer, with the probability $p_{i,j}^-$ as a negative customer, with the probability $q_{i,j}$ as a non-active customer and with the probability $p_{i,0}$ is removed from network ($\sum_{j=1}^N (p_{i,j} + p_{i,j}^- + q_{i,j}) + p_{i,0} = 1$), $i \in J$. Let $p_{i,i} = p_{i,i}^- = q_{i,i} = 0$, $i \in J$.

Let $n_i(t), n_i'(t)$ are numbers of ordinary and non-active customers at system $i \in J$ at time t. Consider $X(t) = \Big((n_1(t), n_1'(t)), \ldots, (n_N(t), n_N'(t))\Big)$. $X(t)$ is a continuous-time Markov chain.

A traffic equations system is:

$$\varepsilon_i = \lambda_i + \sum_{j=1}^N \frac{\varepsilon_j}{\varepsilon_j^- + \mu_j} p_{j,i},$$

$$\varepsilon_i^- = \lambda_i^- + \sum_{j=1}^N \frac{\varepsilon_j}{\varepsilon_j^- + \mu_j} p_{j,i}^-,$$

$$\omega_i = \theta_i + \sum_{j=1}^N \frac{\varepsilon_j}{\varepsilon_j^- + \mu_j} q_{j,i}, \quad i \in J.$$

One can prove that under certain conditions traffic equations system has unique non-trivial solution.

Under the condition of stationary distribution existence stationary probabilities $\pi(n, n')$ ($n = (n_1, \ldots, n_N), n' = (n_1', \ldots, n_N')$) satisfy global balance equations:

$$\pi(n, n')\Big[\sum_{i \in J} \Big(\lambda_i + \mu_i I_{n_i \neq 0} + \lambda_i^- I_{n_i \neq 0} + \nu_i I_{n_i \neq 0} +$$

$$+ \varphi_i I_{n_i' \neq 0} + \theta_i + \psi_i I_{n_i' \neq 0}\Big)\Big] =$$

$$= \sum_{i \in J} \Big[\pi(n - e_i, n')\lambda_i I_{n_i \neq 0} + \pi(n + e_i, n')\mu_i p_{i,0} + \pi(n + e_i, n')\lambda_i^- +$$

$$+\pi(n+e_i,n'-e_i)\nu_i I_{n'_i\neq 0}+\pi(n-e_i,n'+e_i)\varphi_i I_{n_i\neq 0}+$$
$$+\pi(n,n'-e_i)\theta_i I_{n'_i\neq 0}+\pi(n,n'+e_i)\psi_i+$$
$$+\sum_{j\in J}\Big(\pi(n+e_i-e_j,n')\mu_i p_{i,j} I_{n_j\neq 0}+\pi(n+e_i+e_j,n')\mu_i p^-_{i,j}+$$
$$+\pi(n+e_i,n'-e_j)\mu_i q_{i,j} I_{n'_j\neq 0}\Big)\Big].$$

Theorem 2. *Under conditions of ergodicity:*

$$\varepsilon_i<\mu_i+\varepsilon_i^-,$$

$$\omega_i<\psi_i,\ i=1,\dots,N,$$

under conditions of reversibility:

$$\frac{\varepsilon_i}{\varepsilon_i^-+\mu_i}\frac{\nu_i}{\varphi_i}=\frac{\omega_i}{\psi_i},\ i\in J,$$

$X(t)$ has stationary distribution:

$$\pi(n,n')=\pi_1(n_1,n'_1)\pi_2(n_2,n'_2)\dots\pi_N(n_N,n'_N),$$

where

$$\pi_i(n_i,n'_i)=\Big(1-\frac{\varepsilon_i}{\mu_i+\varepsilon_i^-}\Big)\Big(1-\frac{\omega_i}{\psi_i}\Big)\Big(\frac{\varepsilon_i}{\mu_i+\varepsilon_i^-}\Big)^{n_i}\Big(\frac{\omega_i}{\psi_i}\Big)^{n'_i},$$

here ε_i, ε_i^-, ω_i, $i\in J$ – is a solution of a traffic equations system.

REFERENCES

1. *Tsitsiashvili G., Osipova M.* Queueing models with different schemes of customers transformations // Queues: flows, systems, networks (BWWQT - 2007): proceedings of the Int. Conf. "Mathematical Methods for Increasing Efficiency of Information Telecommunication Networks". 2007. № 9. P. 128–133.
2. *Кравченко С. В.* Открытые сети с информационными сигналами // Известия Гомельского государственного университета им. Ф. Скорины. 2002. V. 15. № 6. P. 175–178.
3. *Малинковский Ю. В.* Критерий представимости стационарного распределения состояний открытой марковской сети обслуживания с несколькими классами заявок в форме произведения // Автоматика и телемеханика. 2002. № 4. P. 75–83.
4. *Якубович О. В.* Стационарное распределение сети массового обслуживания с различными типами отрицательных и положительных заявок и ограничением на время ожидания // Известия Гомельского государственного университета им. Ф. Скорины. 2007. V. 45. № 6. P. 198–202.

STOCHASTIC BOUNDS FOR THE MEAN CHARACTERISTICS OF AN M/G/1 QUEUE WITH GENERAL RETRIAL TIMES

M. Boualem[1], N. Djellab[2], D. Aïssani[1]

[1] *Laboratory of Modelization and Optimization of Systems (LAMOS), University of Béjaïa,*

[2] *Department of Mathematics, University of Annaba*

[1] *Béjaïa, Algeria*

[2] *Annaba, Algeria*

robertt15dz@yahoo.fr

The main goal of this paper is to investigate various monotonicity properties of a single server retrial queue with a first-come-first-served (FCFS) orbit and general retrial times using the stochastic order relations (strong stochastic ($\leq_{st}$), increasing convex ($\leq_{icx}$), and Laplace ordering ($\leq_L$)) in order to derive performance indices bounds.

Keywords: Retrial queues, Stochastic ordering, Monotonicity, Ageing distributions.

1. INTRODUCTION

The retrial queueing system has been studied extensively due to its wide applicability. Apart from theoretical interest, it has been successfully applied in telephone switching systems, telecommunication networks, and computer networks [1].
In almost all models of retrial queues, the time between retrials for any customer is assumed to be exponentially distributed. In recent years, retrial queueing systems with general service times and nonexponential retrial time distribution have received little attention [3, 6]. An important characteristic of the general retrial times policy is that we always obtain analytical solutions in terms of closed-form expressions. The general retrial times policy arises naturally in problems where the server is required to search for customers, that is, this policy is related to many service systems where, after each service completion, the processor will spend a random amount of time in order to find the next item to be processed.

Many efforts have been devoted to derive performance measures such as queue length distribution, waiting times distribution, busy period distribution etc. in retrial queues. In many cases the behavior of the retrial queue is described by the Markov chain with spatially inhomogeneous infinitesimal generator (or transition probability matrix) caused by transitions due to repeated attempts. This spatial inhomogeneity often leads the analytical complexity and bounds and/or approximations are used instead.
Stochastic comparison methods have been used to produce bounds and approximations

for queue length processes and waiting times in many queueing systems. For the detailed results about the comparison methods and their applications, e.g. see [4, 5]

In this paper we study some monotonicity properties similar to Boualem et al. [2], for an $M/G/1$ queue with general service times and nonexponential retrial time distribution under FCFS orbit discipline. The performance characteristics of such a system are available in Gómez-Corral [3]. We prove the monotonicity of the transition operator of the embedded Markov chain relative to strong stochastic ordering and convex ordering. We obtain comparability conditions for the distribution of the number of customers in the system. Inequalities are derived for the mean characteristics of the busy period, number of customers served during a busy period, number of orbit busy periods and waiting times.

2. DESCRIPTION OF THE QUEUEING SYSTEM

We consider a single server retrial queue with general service times and nonexponential retrial time distribution under FCFS orbit discipline. Primary customers arrive in a Poisson process with rate λ. If the server is free, the primary customer is served immediately and leaves the system after service completion. Otherwise, the customer leaves the service area and enters the retrial group in accordance with an FCFS discipline. We will assume that only the customer at the head of the orbit is allowed for access to the server. If the server is busy upon retrial, the customer joins the orbit again. Such a process is repeated until the customer finds the server idle and gets the requested service at the time of a retrial. Successive inter-retrial times of any customer follow an arbitrary law with common probability distribution function $A(x)$, Laplace-Stieltjes transform $L_A(s)$ and first moment α_1. The service times are independently and identically distributed with probability distribution function $B(x)$, Laplace-Stieltjes transform $L_B(s)$ and first two moments β_1, β_2. We suppose that inter-arrival times, retrial times and service times are mutually independent. The performance characteristics of such a system are available in Gómez-Corral [3].

Let τ_n be the time of the nth departure and Q_n the number of customers in the orbit just after the time τ_n. We have the following fundamental recursive equation:

$$Q_{n+1} = Q_n + v^{n+1} - \delta_{Q_{n+1}},$$

where v^{n+1} is the number of primary customers arriving at the system during the service time which ends at τ_{n+1}. Its distribution is given by $k_j = \int_0^\infty (\lambda x)^j (j!)^{-1} e^{-\lambda x} dB(x),\ j \geq 0$, with generating function $k(z) = \sum_{j\geq 0} k_j z^j = L_B(\lambda(1-z))$.

The Bernoulli random variable $\delta_{Q_{n+1}}$ is equal to 1 or 0 depending on whether the customer who leaves the system at time τ_{n+1} proceeds from the orbit or otherwise.

The sequence of random variables $\{Q_n\}$ forms an embedded Markov chain for our queueing system which is irreductible and aperiodic on the state-space $\mathbb{N}$. The inequality $\lambda\beta_1 < L_A(\lambda)$ is a necessary and suffcient condition for the system to be stable [3].

2.1. Some useful lemmas. This subsection presents several useful lemmas which will be used later in establishing the main results. Consider two $M/G/1$ retrial queues with classical retrial policy and feedback with parameters $\lambda^{(i)}$, $\theta^{(i)}$ and $B^{(i)}$, $i = 1, 2$. Let $k_j^{(i)}$ be the distribution of the number of primary calls which arrive during the service time of a call in the ith system.

The following two lemmas turns out to be a useful tool for showing the monotonicity properties of the embedded Markov chain.

Lemma 1. *If $\lambda^{(1)} \leq \lambda^{(2)}$ and $B^{(1)} \leq_s B^{(2)}$, then $\{k_n^{(1)}\} \leq_s \{k_n^{(2)}\}$, where $\leq_s$ is one of the symbols $\leq_{st}$ or $\leq_{icx}$.*

Proof. the proof is known in the more general setting of a random summation. □

Lemma 2. *If $\lambda^{(1)} \leq \lambda^{(2)}$ and $B^{(1)} \leq_L B^{(2)}$, then $\{k_n^{(1)}\} \leq_L \{k_n^{(2)}\}$.*

Proof. We have

$$k^{(i)}(z) = \sum_{n \geq 0} k_n^{(i)} z^n = L_{B^{(i)}}(\lambda^{(i)}(1-z)), \ i = 1, 2$$

where $k^{(1)}(z)$, $k^{(2)}(z)$ are the corresponding distributions of the number of new arrivals during a service time.
Let $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_L B^{(2)}$. To prove that $\{k_n^{(1)}\} \leq_L \{k_n^{(2)}\}$, we have to establish that

$$L_{B^{(1)}}(\lambda^{(1)}(1-z)) \geq L_{B^{(2)}}(\lambda^{(2)}(1-z)).$$

□

3. MONOTONICITY PROPERTIES OF THE EMBEDDED MARKOV CHAIN

The one-step transition probabilities of the embedded Markov chain $\{Q_n, n \geq 1\}$ is defined in the following formulae
$p_{nm} = (1 - L_A(\lambda))k_{m-n} + L_A(\lambda)k_{m-n+1}$, for $n \neq 0$ and $m \geq 0$,
$p_{0m} = k_m$, for $m \geq 0$.

Let $\mathbf{T}$ be the transition operator of an embedded Markov chain which associates to every distribution $\omega = \{\omega_m\}_{m \geq 0}$ a distribution $\mathbf{T}\omega = \{\nu_m\}_{m \geq 0}$ such that $\nu_m = \sum_{n \geq 0} \omega_n p_{nm}$. From Stoyan [5], T is monotone with respect to $\leq_{st}$ if and only if

$$\overline{p}_{nm} - \overline{p}_{n-1m} \geq 0 \ \text{ for all } \ n \ \text{ and } \ m,$$

and is monotone with respect to $\leq_v$ if and only if

$$\overline{\overline{p}}_{n-1m} + \overline{\overline{p}}_{n+1m} - 2\overline{\overline{p}}_{nm} \geq 0 \ \text{ for all } \ n \ \text{ and } \ m.$$

Here, $\bar{p}_{n,m} = \sum_{l=m}^{\infty} p_{n,l}$ and $\bar{\bar{p}}_{n,m} = \sum_{l=m}^{\infty} \bar{p}_{n,l}$.

Theorem 1. $\mathbf{T}$ *is monotone with respect to the orders* $\leq_{st}$ *and* $\leq_{icx}$.

Proof. In our case:

$$\begin{aligned}\overline{p}_{nm} &= (1-L_A(\lambda))\overline{k}_{m-n} + L_A(\lambda)\overline{k}_{m-n+1} = (1-L_A(\lambda))k_{m-n} + \overline{k}_{m-n+1},\\ \overline{\overline{p}}_{nm} &= (1-L_A(\lambda))\overline{k}_{m-n} + \overline{\overline{k}}_{m-n+1}.\end{aligned}$$

Thus

$$\begin{aligned}\overline{p}_{nm} - \overline{p}_{n-1m} &= (1-L_A(\lambda))k_{m-n} + L_A(\lambda)k_{m-n+1} \geq 0,\\ \overline{\overline{p}}_{n-1m} + \overline{\overline{p}}_{n+1m} - 2\overline{\overline{p}}_{nm} &= (1-L_A(\lambda))k_{m-n-1} + L_A(\lambda)k_{m-n} \geq 0.\end{aligned}$$

□

In the following two theorems, we give comparability conditions of two transition operators. Consider two $M/G/1$ retrial queues with nonexponential retrial times with parameters $\lambda^{(1)}$, $A^{(1)}$, $B^{(1)}$ and $\lambda^{(2)}$, $A^{(2)}$, $B^{(2)}$ respectively. Let $\mathbf{T}^1$ and $\mathbf{T}^2$ be the transition operators of the corresponding embedded Markov chains.

Theorem 2. *If* $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_s B^{(2)}$ *and* $A^{(1)} \leq_L A^{(2)}$, *then* $\mathbf{T}^1 \leq_s \mathbf{T}^2$, *i.e. for any distribution* ω, *we have* $\mathbf{T}^1\omega \leq_s \mathbf{T}^2\omega$, *where* $\leq_s$ *is one of the symbols* $\leq_{st}$ *or* $\leq_{icx}$.

Proof. From Stoyan [5], it is well known that to prove $\mathbf{T}^1 \leq_s \mathbf{T}^2$, we have to show the following numerical inequalities for the one-step transition probabilities $p^{(1)}_{nm}, p^{(2)}_{nm}$:

$$\overline{p}^{(1)}_{nm} \leq \overline{p}^{(2)}_{nm}, \quad \forall n,\ m, \quad (\text{for } \leq_s = \leq_{st}), \tag{1}$$

$$\overline{\overline{p}}^{(1)}_{nm} \leq \overline{\overline{p}}^{(2)}_{nm}, \quad \forall n,\ m, \quad (\text{ for } \leq_s = \leq_{icx}), \tag{2}$$

To prove inequality (1), we have

$$\overline{p}^{(1)}_{nm} = (1-L_{A^{(1)}}(\lambda^{(1)}))k^{(1)}_{m-n} + \overline{k}^{(1)}_{m-n+1}.$$

Since $\lambda^{(1)} \leq \lambda^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $L_{A^{(1)}}(\lambda^{(1)}) \geq L_{A^{(2)}}(\lambda^{(2)})$ and

$$\overline{p}^{(1)}_{nm} \leq (1-L_{A^{(2)}}(\lambda^{(2)}))k^{(1)}_{m-n} + \overline{k}^{(1)}_{m-n+1}.$$

But

$$(1-L_{A^{(2)}}(\lambda^{(2)}))k^{(1)}_{m-n} + \overline{k}^{(1)}_{m-n+1} = (1-L_{A^{(2)}}(\lambda^{(2)}))\overline{k}^{(1)}_{m-n} + L_{A^{(2)}}(\lambda^{(2)})\overline{k}^{(1)}_{m-n+1}.$$

By Lemma 1, we have $\overline{k}^{(1)}_n \leq \overline{k}^{(2)}_n$, $\forall n \geq 0$.
Using these inequalities we get:

$$\overline{p}^{(1)}_{nm} \leq (1-L_{A^{(2)}}(\lambda^{(2)}))\overline{k}^{(2)}_{m-n} + L_{A^{(2)}}(\lambda^{(2)})\overline{k}^{(2)}_{m-n+1} = \overline{p}^{(2)}_{nm}.$$

Following the technique above it is possible to establish inequality (2). □

Theorem 3. *If* $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_L B^{(2)}$ *and* $A^{(1)} \leq_L A^{(2)}$, *then* $\mathbf{T}^1 \leq_L \mathbf{T}^2$.

Proof. Let $\omega = (\omega_m)$ be a distribution and $\mathbf{T}_\omega = \nu = (\nu_m)$, where
$\nu_m = \sum_{n\geq 0} \omega_n p_{nm} = \omega_0 k_m + \sum_{n\geq 1} \omega_n p_{nm}$, for all $m \geq 0$.
Let $k(z) = \sum_{n\geq 0} k_n z^n$ and $\omega(z) = \sum_{n\geq 0} \omega_n z^n$ be the generating functions of (k_n) and (ω_n) respectively. The generating function of ν is given by

$$\begin{aligned} G(z) &= \sum_{m\geq 0} \nu_m z^m = \sum_{m\geq 0}\sum_{n\geq 0} \omega_n p_{nm} z^m = \sum_{m\geq 0}[\omega_0 k_m + \sum_{n\geq 1} \omega_n p_{nm}] z^m \\ &= \omega_0 k(z) + \frac{1}{z} k(z)(\omega(z) - \omega_0)(z + (1-z)L_A(\lambda)). \end{aligned}$$

If the conditions of Theorem 3 are fulfilled, then $k^{(1)}(z) \geq k^{(2)}(z)$ by Lemma 2 and $(1-z)L_{A^{(1)}}(\lambda^{(1)}) \geq (1-z)L_{A^{(2)}}(\lambda^{(2)})$, $\forall\, z \in [0,1]$. Hence $G^{(1)}(z) \geq G^{(2)}(z)$. □

4. BOUNDS FOR THE MEAN CHARACTERISTICS OF THE SYSTEM

The main characteristics of a system busy period, the orbit busy period and waiting time are:
L: the length of a system busy period,
I: the number of service completions occurring during $(0, L]$,
N_b: the number of orbit busy periods which take place in $(0, L]$,
W: the waiting time.
Gómez-Corral [3] shows that, if $\lambda\beta_1 < L_A(\lambda)$, then
$E(L) = \frac{\beta_1}{L_A(\lambda)-\lambda\beta_1}$, $E(I) = \frac{L_A(\lambda)}{L_A(\lambda)-\lambda\beta_1}$, $E(N_b) = \frac{1-L_B(\lambda)}{L_B(\lambda)}$, and $E(W) = \frac{\lambda\beta_2 + 2\beta_1(1-L_A(\lambda))}{2(L_A(\lambda)-\lambda\beta_1)}$.

Suppose once more that we have two $M/G/1$ retrial queues with nonexponential retrial times with parameters $\lambda^{(1)}$, $A^{(1)}$, $B^{(1)}$ and $\lambda^{(2)}$, $A^{(2)}$, $B^{(2)}$, respectively. Let $L^{(i)}$, $I^{(i)}$, $N_b^{(i)}$ and $W^{(i)}$ be the length busy period, the number of customers served during a busy period, the number of orbit busy periods which take place in $(0, L^{(i)}]$ and the waiting time respectively, in the i-th system, $i = 1, 2$.

Theorem 4. *If $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_s B^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $E(L^{(1)}) \leq E(L^{(2)})$, and $E(I^{(1)}) \leq E(I^{(2)})$, where $\leq_s$ is one of the symbols $\leq_{st}$, $\leq_{icx}$, $\leq_L$.*

Proof. The quantities $E(L)$ and $E(I)$ which are increasing with respect to λ and β_1, decreasing with respect to $L_A(.)$. Under conditions of Theorem 4, we obtain the desired inequalities. Recall that $X \leq_s Y$ implies $E(X^n) \leq E(Y^n)$ for all n. □

Theorem 5. *For any $M/G/1$ retrial queue,*
$E(L) \leq \frac{\beta_1}{e^{-\lambda\alpha_1}-\lambda\beta_1}$, *and* $E(I) \leq \frac{e^{-\lambda\alpha_1}}{e^{-\lambda\alpha_1}-\lambda\beta_1}$.
If A and B are $\mathcal{L}$, then $E(L) \geq \frac{\beta_1(1+\lambda\alpha_1)}{1-\lambda\beta_1(1+\lambda\alpha_1)}$, *and* $E(I) \geq \frac{1}{1-\lambda\beta_1(1+\lambda\alpha_1)}$.

Proof. We consider auxiliary $M/D/1$ and $M/M/1$ retrial queues with the same arrival rates λ, mean service times β_1 and mean retrial times α_1. A is Dirac distribution at α_1 for the $M/D/1$ system, and is exponential distribution for the $M/M/1$ system. Using the theorem above we obtain the stated results. □

Theorem 6. *If* $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_{st} B^{(2)}$ *and* $A^{(1)} \leq_L A^{(2)}$, *then*

$$E(N_b^{(1)}) \leq E(N_b^{(2)}), \text{ and } E(W^{(1)}) \leq E(W^{(2)}).$$

Proof. The quantities $E(N_b)$ and $E(W)$ are increasing with respect to λ, β_1 and β_2, decreasing with respect to $L_B(.)$ and $L_A(.)$. Under the conditions of Theorem 6 we obtain the desired inequalities. □

Theorem 7. *For any* $M/G/1$ *retrial queue,*
$E(N_b) \leq e^{\lambda\beta_1} - 1$, *and* $E(W) \leq \frac{\lambda\beta_2 + 2\beta_1(1-e^{-\lambda\alpha_1})}{2(e^{-\lambda\alpha_1} - \lambda\beta_1)}$.
If B *and* A *are* $\mathcal{L}$, *then*
$E(N_b) \geq \lambda\beta_1$, *and* $\frac{\lambda\beta_2(1+\lambda\alpha_1)+2\lambda\beta_1\alpha_1}{2(1-\lambda\beta_1(1+\lambda\alpha_1))} \leq E(W) \leq \frac{2\lambda\beta_1^2+2\beta_1(1-e^{-\lambda\alpha_1})}{2(e^{-\lambda\alpha_1}-\lambda\beta_1)}$.

Proof. The proof is similar to that of Theorem 5. In addition, if a given distribution F is $\mathcal{L}$ then $\beta_2 \leq 2\beta_1^2$. □

REFERENCES

1. *Artalejo J.R. , Gómez-Corral A.* Retrial queueing system: A computation approach // Springer Edition, Berlin, 2008.
2. *Boualem M., Djellab N., Aïssani D.* Stochastic inequalities for $M/G/1$ retrial queues with vacations and constant retrial policy // Mathematical and Computer Modelling. 2009. V. 50. № 1-2. P. 207–212.
3. *Gómez-Corral A.* Stochastic analysis of single server retrial queue with the general retrial times // Naval Research Logistics. 1999. V. 46. P. 561–581.
4. *Shaked M., Shanthikumar J. G.* Stochastic Orders // Springer-Verlag, New York, 2007.
5. *Stoyan D.* Comparison methods for queues and other stochastic models // Wiley, New York, 1983.
6. *Taleb S., Saggou H., Aissani A.* Unreliable M/G/1 retrial queue with geometric loss and random reserved time // Int. J. Operational Research. 2010. V. 7. № 2. P. 171–191.

HEAVY TRAFFIC ANALYSIS OF M/G/1 QUEUE WITH MARKOV MODULATED SERVICE SPEED

M. Dimitrov, M. Ivanova
UNWE
Sofia, Bulgaria
mcdimitrov@abv.bg

An M/G/1I queue whose service speed changes according to an external environment that is governed by a Markov process is considered. The server could change its service speed many times while serving a customer. The asymptotic behavior of the virtual workload and its mean value is investigated under heavy traffic assumptions.

Keywords: single server queue,Markov - modulated service speed, heavy traffic.

1. INTRODUCTION

A single server queueing system with infinite waiting space where customers bring a random amount of work is considered. The server processes the work of different speeds. For example, in computer processors, web servers, a single processor, works on multiple jobs in parallel of which we are interested in our job performance. In this case the processor could be modeled as a single server with processing speeds that vary according to a stochastic environment process. On the whole, there are much more investigations of queue with time varying and Markov - modulated arrivals, than on time varying of the service speed, for example [1], [2], [3].

In this paper, a queue is studied whose service capacity varies over time. That is, the speed of the server with which it serves a customer, is determined by an external environment process. We assume that the server speed changes according to a continuous time Markov chain that is independent of the arrival process and service requirements of the customer. Each customer brings a certain random amount of work and the rate at which this work is completed is time varying. We assume that the customer in the queue are served in a First Come First Served manner. For this model we obtain stationary distribution of workload in heavy traffic assumptions.

S.R. Mahabhashyam and N. Gautam [3] considered a queue with Markov - modulated speed and exponentially distributed workload. We'll consider M/G/1 queueing model with Markov modulated service speed as in [4], pp. 224-226.

2. PROBLEM DESCRIPTION

Customer arrive into the queue according to a Poisson process with mean rate λ customers per unit time. Each arriving customer brings a random amount of work with

distribution function $H(x)$, Laplace - Stieltjes transform $h(s)$ and first two moments h_1 and h_2. Let $Z(t)$ be the state of the environmental process which governs the server speed at time t such that $Z(t)$ is an continuous time markov chain with finite state space {0,1,2,...,K}. That is, the server can do E_i amount of work per unit time. Let α_{nm} be rates of transitions of the process $Z(t)$, π_n, $n = 0, 1, \ldots, K$ be its stationary distribution, $\alpha_n = -\alpha_{nn}$, $Q = \|\alpha_{ij}\|$. Arriving customers join unlimited queue and then are served in accordance with FIFO discipline.

The system is ergodic if $\lambda h_1/\overline{E} < 1$ and $\bar{E} = \sum\limits_{i=0}^{K} \pi_i E_i$. From now on we assume that the inequality $\lambda h_1/\bar{E} < 1$ holds and the system is in steady state.

The virtual workload process $W(t)$ at time t is considered. The asymptotic behavior of $W(t)$ and $W = EW(t)$ under heavy traffic, when $\rho = \lambda h_1/\overline{E} \to 1-0$ in steady state is investigated.

The set of equations $Q(x_o, x_1, \ldots, x_K)^T = (b_0, b_1, \ldots, b_K)^T$ has a solution if the vector $(b_0, b_1, \ldots, b_K)$ is orthogonal to the vector $\overline{\pi}$.

Then set of linear algebraic equations

$$\sum_{j=0}^{K} \alpha_{ji} x_j = (E_i - \overline{E}), 0 \le i \le K \tag{1}$$

has a solution.

From lemma 2.3 Falin [2], follows, that for any vector $\overline{y} = (y_0, y_1, \ldots, y_K)$ we have

$$\overline{y}QR = \overline{y} - \frac{y_0}{\pi_0 \pi}, \bar{X} = R(E_0 - \bar{E}, \ldots, E_K - \bar{E})^T \text{where} R = \begin{pmatrix} 0 & \ldots & \ldots & 0 \\ 0 & & & \\ \vdots & & A^{-1} & \\ 0 & & & \end{pmatrix} \tag{2}$$

and the matrix $A = \|\alpha_{ij}\|, 1 \le i, j \le K$

3. THE MEAN VIRTUAL WORKLOAD

Theorem 1. *The mean virtual workload is equal to*

$$W = \frac{\frac{\lambda h_2}{2} + \sum\limits_{i=0}^{K} x_i \pi_i (\lambda h_1 - E_i) + \sum\limits_{i=0}^{K} x_i a_i E_i}{\overline{E}(1-\rho)} \tag{3}$$

and

$$\sum_{i=0}^{K} a_i E_i = \overline{E}(1-\rho) \tag{4}$$

Proof. We consider two - dimensional Markov process $\{W(t), Z(t)\}$. Using Markov property of the process the next system of integro - differential equations follows

$$E_i F_i'(x) - (\lambda + \alpha_n) F_i(t) + \lambda \int_0^x H(x-y) dF_i(y) + \sum_{j \ne i} \alpha_{ji} F_j(x) = 0, (0 \le i \le K) \tag{5}$$

Let $a_n = P(W(t) = 0, Z(t) = n)$, $\varphi_n(s) = E(e^{-sW(t)}; Z(t) = n) = a_n + \int\limits_{0+}^{\infty} e^{-sx} dF_n(x)$

Equation (5) in Laplace - Stieltjes transforms φ_n is

$$\sum_{j=0}^{K} \varphi_j(s)\alpha_{ji} = [\lambda(1 - h(s)) - sE_i]\varphi_i(s) + sE_i a_i, 0 \le i \le K \tag{6}$$

We'll use the property of the matrix Q, and summing the equations (6) on i we get

$$\sum_{i=0}^{K} a_i E_i = \sum_{i=0}^{K} \left[E_i - \lambda \frac{1 - h(s)}{s}\right] \varphi_i(s) \tag{7}$$

As $s \to 0$ from (7) follows (4).

Really $\sum\limits_{i=0}^{K} E_i a_i = \sum\limits_{i=0}^{K} [E_i - \lambda h_1]\pi_i = \overline{E} - \lambda h_1 = \overline{E}(1 - \rho)$.

Differentiating both sides of 7 at $s = 0$ we have

$$\sum_{i=0}^{K} (E_i - \lambda h_1) W_i = \frac{\lambda h_2}{2} \tag{8}$$

To find $W = \sum\limits_{i=0}^{n} W_i$ we will multiply both sides of (6) by x_i and sum with respect to i. Then from $\sum\limits_{i} \alpha_{ji} x_i = E_j - \overline{E}$ we get

$$\sum_{j=0}^{K} (E_j - \overline{E})\varphi_j(s) = s \sum_{i=0}^{K} x_i [\lambda \frac{(1 - h(s))}{s} - E_i]\varphi_i(s) + \sum_{i=0}^{K} x_i E_i a_i s$$

Taking first derivative from this equation with respect to s at $s = 0$ we have

$$-\sum_{j=0}^{K} (E_j - \overline{E}) W_j = \sum_{i=0}^{K} x_i \pi_i (\lambda h_1 - E_i) + \sum_{i=0}^{K} x_i E_i a_i \tag{9}$$

Summing (8) and (9) we get the mean value W in (3) □

Theorem 2. *Under heavy traffic*

$$\overline{E}(1 - \rho) W = \frac{h_2}{2h_1} + \sum_{i=0}^{K} x_i \pi_i (\overline{E} - E_i) + \circ(1) \tag{10}$$

Proof. From $\sum\limits_{i=0}^{K} a_i E_i = \overline{E}(1 - \rho)$, and $\overline{E} > 0$ when $\rho \to 1 - 0$ follows $a_i \to 0$ for all $i = 0, 1, 2, \ldots, K$.

From (4) and $\sum\limits_{i=0}^{K} x_i a_i E_i = \circ(1)$ follows (10) □

Theorem 3. *Let $x_* = min\{x_i\}$, $x^* = max\{x_i\}$. Then*

$$x_* \leq W - \frac{\frac{\lambda h_2}{2} + \sum\limits_{i=0}^{K} x_i \pi_i (\lambda h_i - E_i)}{\overline{E}(1-\rho)} \leq x^* \tag{11}$$

Proof. Since $a_i > 0$ for all $i = 0, 1, 2, \ldots, K$, then

$$\sum_{i=0}^{K} x_i a_i E_i \geq x_* \sum_{i=0}^{K} a_i E_i = x_* \overline{E}(1-\rho)$$

and

$$\sum_{i=0}^{K} x_i a_i E_i \leq x^* \sum_{i=0}^{K} a_i - E_i = x^* \overline{E}(1-\rho)$$

or

$$x_* \leq \frac{\sum\limits_{i=0}^{K} x_i a_i E_i}{\overline{E}(1-\rho)} \leq x^*.$$

From (3) follows (11). □

Theorem 4. *Under heavy traffic*
a)random variables $W(t)$ and $Z(t)$ are asymptotically independent;
b)the random variable $\overline{E}(1-\rho)W(t)$ is asymptotically exponential with the mean

$$\frac{\overline{E} h_2}{2h_1} + \sum_{i=0}^{K} \pi_i x_i (\overline{E} - E_i) \tag{12}$$

Proof. From (6) follows the next equation

$$\overline{\varphi}(s) Q = \overline{\varphi}(s)[\lambda(1-h(s))E - s\tilde{E}] + s\bar{a} \tag{13}$$

where $\overline{\varphi} = (\varphi_0(s), \varphi_1(s), \ldots, \varphi_K(s))$, $\bar{a} = (a_o^{E_0}, a_1^{E_1}, \ldots, a_K^{E_K})$, $E = diag(1, 1, \ldots, 1)$, $\tilde{E} = diag(E_0, E_1, \ldots, E_K)$. From $Q\bar{e} = 0$, $\bar{a}\bar{e}^T = \sum\limits_{i=0}^{K} a_i E_i = \overline{E}(1-\rho) = \varepsilon$, multiplying (13) by $\bar{e}^T$ we get

$$\bar{\varphi}(s)[\lambda(1-h(s))E - s\tilde{E}]\bar{e}^T + s\varepsilon = 0 \tag{14}$$

Again, multiplying (13) by R and from $\bar{x}QR = \bar{x} - x_0/\pi_0\bar{\pi}$ we get

$$\bar{\varphi}(s) = \frac{\varphi_0(s)\bar{\pi}}{\pi_0} + \bar{\varphi}(s)[\lambda(1-h(s))E - s\tilde{E}]R + s\bar{a}R \tag{15}$$

Substituting $\bar{\varphi}(s)$ from 15 into the right side of 15 we have

$$\varphi(s) = \frac{\varphi_0(s)}{\pi_0}\overline{\pi}[E + \lambda(1-h(s))R - s\tilde{E}R] + sY(s) \tag{16}$$

where $Y(s) = s\varphi(s)[\lambda(1-h(s))R - s\tilde{E}R]^2 + \bar{a}R[E + \lambda(1-h(s))R - s\tilde{E}R]$.
Now replacing $\varphi(s)$ in 14 by $\varphi(s)$ from 16 we get.

$$\varphi_0(s) = -\pi_0 \frac{A}{B} \tag{17}$$

where $A = sY(s)[\lambda(1-h(s))E - s\tilde{E}]\bar{e}^T + s\varepsilon$ and

$$B = B_1 + B_2 = \bar{\pi}[\lambda(1-h(s))E - s\tilde{E}]\bar{e}^T + \bar{\pi}[\lambda(1-h(s))R - s\tilde{E}R][\lambda(1-h(s))E - s\tilde{E}]\bar{e}^T$$

.

Finally, let‘s replace s by εs and λ increases in such a way that $\varepsilon = \bar{E}(1-\rho) \to 0$.
From $|\varphi_i(\varepsilon s)| \leq 1$ and $\bar{a}\bar{e}^T = \bar{E}(1-\rho) \to 0$ follows $\bar{Y}(\varepsilon s) \to 0$ as $\varepsilon \to 0$, also $\frac{h(\varepsilon s)-1+h_1\varepsilon s}{\varepsilon^2 s^2} \to \frac{h_2}{2}$, $\frac{1-h(\varepsilon s)}{\varepsilon s} \to h_1$.
Now $B_1(\varepsilon s) = -\lambda[h(\varepsilon s) - 1 + h_1(\varepsilon s)] - s\bar{E}(1 - \frac{\lambda h_1}{\bar{E}}) = -\varepsilon^2 s^2 \lambda \frac{[h(\varepsilon s)-1+h_1 s\varepsilon]}{\varepsilon^2 s^2} - s\varepsilon^2$ and $-B_2(\varepsilon s) = -\pi s^2 \varepsilon^2 [\lambda \frac{(1-h(\varepsilon s))}{\varepsilon s} - \tilde{E}]R[\lambda \frac{1-h(\varepsilon s)}{\varepsilon s} E - \tilde{E}]\bar{e}^T$.
Thus $\lim\limits_{\varepsilon \to 0} \varphi_0(\varepsilon) = \frac{\pi_0}{1+MS}$, where $M = \frac{\overline{E}h_2}{2h_1} - \pi(\bar{E}E - \tilde{E})R[\bar{E}E - \tilde{E}]\bar{e}^T$.
Since $R[\bar{E}E - \tilde{E}]\bar{e}^T = -\bar{x}$, then $M = \frac{E\bar{h}_2}{2h_1} + \sum \pi_i x_i(\bar{E} - E_i)$.
Now, from (16) follows *Theorem 4*.
Really $\lim\limits_{\varepsilon \to 0} \varphi(\varepsilon s) = \lim\limits_{\varepsilon \to 0} \varphi_0(\varepsilon s)\frac{\bar{\pi}}{\pi_0} = \pi \frac{1}{1+Ms}$. □

REFERENCES

1. *Prabhulin N. U., Y. Zhu.* Markov - modulated queueing systems // Thechnical report . 1989. V. 2. № 835. P. 1–49.
2. *Falin G. I., Falin A. I.* Heavy traffic analysis of M/G/1 type queueing systems with Markov-modulated arrivals // TOP. 1999. V. 7. № 2. P. 279–291.
3. *Mahabhashyam S. R., Gautam N. O.* On queues with Markov modulated service rates // Queueing systems, V. 51. № 2. P. 89–113.
4. *Гнеденко Б. В. , Коваленко И. Н.* Введение в теорию массового обслуживания // Издательство ЛКИ. 2001. P. 224–226.

ON THE $M^X/G/1$ RETRIAL QUEUE WITH IMPATIENT CUSTOMERS

N.V. Djellab[1], K.N. Arrar[1], J.B. Baillon[2]

[1] *University of Annaba*

[2] *University Paris 1 Pantheon Sorbone*

[1] *Annaba, Algeria*

[2] *UFR 27, CERMSEM CNRS-UMR 8095, France*

djellab@yahoo.fr

In this work, we carry out a stochastic analysis of the $M/G/1$ retrial queue with batch arrivals and impatient customers. Our study includes: steady state joint distribution of the server state and the number of customers in retrial group, embedded Markov chain and stochastic decomposition for the number of customers in the system. We investigate also the asymptotic behaviour of the system under high and low retrial intensities.

Keywords: retrial queue, batch arrival, impatient customer, decomposition property, low retrial rate, convergence.

1. INTRODUCTION

In telephone networks, we can observe that a calling subscriber after some unsuccessful retrials gives up further repetitions and abandons the system. In queueing systems with repeated attempts (retrial queues), this phenomenon is represented by the set of probabilities $\{H_k,\ k \geq 1\}$, called the persistence function, where H_k is the probability that after the kth attempt fails, a customer will make the $(k+1)$th one. In general, it is assumed that the probability of a customer reinitiating after failure of a repeated attempt does not depend on the number of previous attempts (i.e. $H_2 = H_3 = ...$) [2]. In this paper, we consider a single server queueing system at which primary customers arrive in batches of size k (with probability $c_k(t),\ \ k \geq 1$) according to a Poisson stream with rate $\lambda > 0$. If the server is busy at the arrival epoch, then with probability $1 - H_1$ all these customers leave the system without service and with probability $0 < H_1 < 1$ join the retrial group (orbit); whereas if the server is idle, then one of the arriving customers begins his service and leaves the system after the service and the others go to the orbit. Let $c_k = \lim\limits_{k\to\infty} c_k(t),\ k \geq 1$. Thus, $C(z) = \sum\limits_{k=1}^{\infty} c_k z^k$ is the generating function of the steady state distribution of the batch size and $\overline{c} = C^{'}(1)$ is the mean batch size. The impatience phenomenon is represented by the probability $1 - H_1$ (the probability $H_2 = 1$). Any orbiting customer will repeatedly retry until the time at which he finds the server idle and starts his service. The retrial times are exponentially distributed with distribution function $T(x) = 1 - e^{-\theta x}, x \geq 0$, having finite mean $\frac{1}{\theta}$. The service

times follow a general distribution with distribution function $B(x)$ and Laplace-Stieltjes transform $\widetilde{B}(s), Re(s) > 0$. Let $\beta_k = (-1)^k \widetilde{B}^{(k)}(0)$ be the kth moment of the service time about the origin and $\rho = \lambda H_1 \beta_1 \overline{c}$ be the traffic intensity. Finally, we admit the hypothesis of mutual independence between all random variables defined above.

2. STEADY STATE DISTRIBUTION OF THE SYSTEM STATE AND DECOMPOSITION PROPERTY

The state of the considered system at time t can be described by means of the process $\{C(t), N_o(t), \zeta(t), t \geq 0\}$, where $N_o(t)$ is the number of customers in the orbit, $C(t)$ is the state of the server at time t. We have that $C(t)$ is 0 or 1 depending on whether the server is idle or busy. If $C(t) = 1$, $\zeta(t)$ represents the elapsed service time of the customer in service at time t.
Let $\rho < 1$. Define

$$\begin{aligned} p_{0n} &= \lim_{t\to\infty} P(C(t) = 0, N_o(t) = n), \\ p_{1n} &= \int_0^\infty \left[\lim_{t\to\infty} \frac{d}{dx} P(C(t) = 1, \zeta(t) \leq x, N_o(t) = n) \right] dx. \end{aligned}$$

In the previous work [1], we found the partial generating functions

$$\begin{aligned} P_0(z) &= \sum_{n=0}^{\infty} z^n p_{0n} = \frac{H_1(1-\rho)}{\rho + H_1(1-\rho)} \mathbf{e}^{\left[\frac{\lambda}{\theta} \int_1^z \frac{1-\widetilde{B}(\lambda H_1(1-C(u)))\frac{C(u)}{u}}{\widetilde{B}(\lambda H_1(1-C(u)))-u} du\right]}, \\ P_1(z) &= \sum_{n=0}^{\infty} z^n p_{0n} = \frac{1-\widetilde{B}(\lambda H_1(1-C(z)))}{(\widetilde{B}(\lambda H_1(1-C(z))) - z)H_1} P_0(z). \end{aligned}$$

We also investigated the steady state queue size distribution at departure epochs ξ_k (the time when the server enters the idle state for the k-th time). The obtained generating function of the steady state distribution of the embedded Markov chain has the following form:

$$\varphi(z) = \frac{1-\rho}{\overline{c}\rho} \frac{\widetilde{B}(\lambda H_1(1-C(z)))(1-C(z))}{\widetilde{B}(\lambda H_1(1-C(z))) - z} \mathbf{e}^{\left[\frac{\lambda}{\theta} \int_1^z \frac{1-\widetilde{B}(\lambda H_1(1-C(u)))\frac{C(u)}{u}}{\widetilde{B}(\lambda H_1(1-C(u)))-u} du\right]}. \quad (1)$$

At present, we establish that the right hand part of (1) can be decomposed into two factors: the first factor $\frac{1-\rho}{\overline{c}\rho} \frac{\widetilde{B}(\lambda H_1(1-C(z)))(1-C(z))}{\widetilde{B}(\lambda H_1(1-C(z)))-z}$ is the generating function for the number of customers at departure epochs in the ordinary $M^X/G/1$ queue with impatient customers (model M_∞), and the second one $\mathbf{e}^{\left[\frac{\lambda}{\theta} \int_1^z \frac{1-\widetilde{B}(\lambda H_1(1-C(u)))\frac{C(u)}{u}}{\widetilde{B}(\lambda H_1(1-C(u)))-u} du\right]} = \frac{P_0(z)}{P_0(1)}$ represents the generating function for the number of customers at the departure epochs in the corresponding retrial queue (model M_θ) given that the server is idle. Thus, stochastic decomposition property of the considered system can be expressed in the following manner:

$$\{0, N_{o\theta}(t), t \geq 0\} = \{0, N_{q\infty}(t), t \geq 0\} + \{0, R_\theta(t), t \geq 0\}.$$

The processes $\{0, N_{o\theta}(t), t \geq 0\}$ and $\{0, R_\theta(t), t \geq 0\}$ are related to the model M_θ, where $R_\theta(t)$ represents the number of customers in the orbit at time t given that the server is idle. The process $\{0, N_{q\infty}(t), t \geq 0\}$ is associated with the model M_∞, where $N_{q\infty}(t)$ is the number of customers in the waiting line at time t.

3. ASYMPTOTIC BEHAVIOUR OF THE SYSTEM

Now, we investigate the asymptotic behaviour of the system under low retrial rate ($\theta \to 0$). To this end, the following result is established:

Theorem 1. *If $\beta_2 < \infty$, then as $\theta \to 0$, the number of customers in the orbit is asymptotically Gaussian with mean $\frac{\lambda}{\theta}\frac{\overline{c}+\rho-1}{1-\rho}$ and variance $\frac{\lambda}{\theta}\frac{\beta_1+\rho\overline{c}\beta_1^2+C''(1)\beta_1^2-\beta_1\rho^3-\rho^2\overline{c}\beta_1^2+\overline{c}\rho^2\beta_2}{2(1-\rho)^2\beta_1^2}$.*

In general, under high retrial intensity ($\theta \to \infty$) the steady state distribution of a retrial queue converges to a limit, which is usually the steady state distribution of a certain limit system. In our case, it is intuitive that this is the model M_∞. We prove this heuristic argument with the help of the stochastic decomposition property.
Let

$$\begin{aligned}
\pi_n(\theta) &= \lim_{k\to\infty} P(N_{o\theta}(\xi_k) = n),\\
\pi_n(\infty) &= \lim_{k\to\infty} P(N_{q\infty}(\xi_k) = n),\\
q_n(\theta) &= \lim_{k\to\infty} P(R_\theta(\xi_k) = n).
\end{aligned}$$

Theorem 2. *The following inequalities take place*

$$2\frac{1-\rho}{\overline{c}\rho\widetilde{B}(\lambda H_1)}(1-q_0(\theta)) \leq \sum_{n=0}^{\infty}|\pi_n(\theta)-\pi_n(\infty)| \leq 2(1-q_0(\theta)), \quad \textit{where}$$

$$q_0(\theta) = \mathbf{e}^{\left[\frac{\lambda}{\theta}\int_1^z \frac{1-\widetilde{B}(\lambda H_1(1-C(u)))\frac{C(u)}{u}}{\widetilde{B}(\lambda H_1(1-C(u)))-u}du\right]} \quad \textit{and} \quad \rho=\lambda H_1\beta_1\overline{c}.$$

$$As \quad \theta \to \infty, \quad \sum_{n=0}^{\infty}|\pi_n(\theta)-\pi_n(\infty)| = 0\left(\frac{1}{\theta}\right).$$

REFERENCES

1. *Djellab N., Arrar K. N., Baillon J. B.* Asymptotic behaviour of the number of customers in retrial group of M/G/1 retrial queue with batch arrivals // International Conference Modern Stochastics: Theory and Applications II, Kyev, Ukraine. 2010. P. 7–11.
2. *Falin G. I., Templeton J. G. C.* Retrial Queues // Chapman and Hall 1997.

A CLOSED SINGLE-LINE QUEUEING SYSTEM WITH REPEATED CALLS AND LOSSES

V. Dragieva
University of Forestry
Sofia, Bulgaria
vildrag2001@yahoo.com

The closed single-server queueing model with repeated attempts is studied in a number of papers. In all these papers it is assumed that a rejected subscriber repeats his attempts until he is satisfied. In the present paper we consider the same model, assuming some restrictions on the number of retrials, which the customer is inclined (or is allowed) to make. The theoretical study of these models being too hard, we investigate them via simulations.

Keywords: Closed queueing systems, repeated calls, simulations

1. INTRODUCTION AND MATHEMATICAL MODEL

The queuing systems with repeated attempts (or retrials) are widely used to model problems in telephone, computer, communication systems, etc. The most obvious example of a retrial queue appears in a telephone call, when a person phones and finds the line engaged. Usually in such a situation the subscriber repeats his attempt after some time. A description of other situations in which retrials arise can be found in [5], and more recent applications of retrial models - in [1]. The standard assumption in analysis of retrial queues is that each customer repeats his attempts until he is satisfied. However, in many real situations this assumption does not hold. It does not always hold even in the classical example of the telephone subscriber who finds engaged line. The objective of the present paper is investigation of retrial queueing models, where the number of repetitions is in some way restricted. The theoretical study of such systems is too difficult, so in the present paper a method of simulations is applied.

We consider retrial queueing system of type $M/G/1//N$ in Kendall's notation. This means that the system has one server, which serves $N, (2 \leq N < \infty)$ customers (subscribers) and each one of the customers produces a Poisson process of demands (calls) with intensity λ. These customers are identified as primary or free customers or sources of primary calls. When an arriving primary customer finds the server busy, he leaves the service area and repeats the request after some random time. Between trials a customer is said to be in orbit (queue), or to be a source of repeated calls, or an active customer. Thus, when a source is free at time moment t (i.e. is not being served and is not waiting for service) it may generate a primary call during the interval $(t,\ t+dt)$ with probability λdt. If the server is free at the instant of arrival of a primary call then

the call starts to be served. During the service time the source cannot generate a new primary call. After service the source moves into a free state and can generate a new primary call. If the channel is busy at time of the arrival of a primary call, then the source starts generating a Poisson flow of repeated calls with intensity μ. As in the case of a primary call, after service the source becomes free and can generate a new primary call. We assume that primary calls, repeated attempts and service times are mutually independent and denote the service time distribution function by $G(x)$ both for primary and repeated calls and its first moment by ν^{-1}.

This queueing system and its variants are useful in modeling magnetic disk memory systems [9], a star-like local area networks [6], [8], telephone networks [7], etc and is studied in a number of papers ([3], [4], [7], [9] and others). All these papers deal only with the case, when each secondary customer repeats his retrials until he is satisfied. Now we consider two variants of the described system, regarding some restrictions on the number of retrials, that a secondary subscriber is inclined (or is allowed) to make before he reaches the server free. In the first variant we assume that if the server is busy at the instant of arrival of a secondary subscriber (an active subscriber), this subscriber continues his repeated attempts with probability p and with probability $1-p$, $0 \leq p \leq 1$ he refuses to try entering server any more and becomes again a free one. In the second model we assume, that each secondary customer stops his retrials if his $(k+1)$-th attempt fails, $k = 1, 2, \ldots$ (if not served before). Further in this paper we will call each of these three models as the model without restrictions (unrestricted model or a model without losses), the first restricted model and the second restricted model, respectively. Obviously, if in the first restricted model we take $p = 1$, or in the second one $k = \infty$ we get the model without restrictions. In the same way, if $p = 0$ the first restricted model coincides with the second one with $k = 1$.

2. ANALYSIS OF THE RESTRICTED MODELS

Let $C(t)$ be the number of busy channels at time t (i.e. $C(t)$ is 0 or 1 according to whether the channel is free or busy at time t) and $A(t)$ - the number of active customers at time t (orbit size or queue length). In the first restricted model (where a rejected secondary subscriber continues his attempts with probability p and with probability $1-p$ stops the retrials) the dynamics of the system can be described by the process $(C(t),\ A(t))$. It should be noted that the situation $C(t) = 0$, $A(t) = N$ is impossible and thus the state space of the process $(C(t),\ A(t))$ is the set $\{0, 1\} \times \{0, 1, \ldots, N-1\}$. As $(C(t),\ A(t))$ is a Markov process only when $C(t) = 0$, in order to work with a Markov process also in the case $C(t) = 1$ we introduce a supplementary variable $z(t)$, equal to the elapsed service time. Let

$$p_{1j}(x)dx = \lim_{t\to\infty} P(C(t) = 1,\ A(t) = j,\ x < z(t) \leq x + dx),$$

$$p_{ij} = \lim_{t\to\infty} P(C(t) = i, A(t) = j),\ \ i = 0, 1,\ \ \ j = 0, 1, \ldots, N-1. \tag{1}$$

In a general way we obtain the equations of statistical equilibrium

$$\frac{d}{dx}p_{1j}(x) = -[(N-j-1)\lambda + \eta(x) + j(1-p)\mu]p_{1j}(x) + (N-j)\lambda p_{1,j-1}(x) \qquad (2)$$

$$+(j+1)(1-p)\mu p_{1,j+1}(x),\ p_{1,-1}(x) = p_{1N}(x) = 0,$$

$$[j\mu + (N-j)\lambda]p_{0j} = \int_0^\infty p_{1j}(x)\eta(x)dx, \qquad (3)$$

$$p_{1j}(0) = (N-j)\lambda p_{0j} + (j+1)\mu p_{0,j+1}, \quad j = 0, 1, \ldots, N-1, \qquad (4)$$

Here

$$\eta(x) = \frac{G'(x)}{1-G(x)}$$

is the service rate at instant x after start of a service.

Recurrent formulas for computing the solutions of the system (2) - (4) in case $p = 1$ are obtained in [9] with the help of discrete transformations and in [3] directly, by means of mathematical induction. For arbitrary p, no formulas are known to us for the solution of this system.

As described before, in the second restricted model we assume that if the $(k+1)$-th attempt of an active subscriber fails, $k = 1, 2, \ldots$, the subscriber leaves the orbit and becomes free again. So, the state of the system at time t can be described by the process $(C(t),\ A_1(t),\ A_2(t), \ldots, A_k(t))$, where $A_i(t)$ is the number of those active subscribers in the orbit which have made exactly k attempts before the moment t, $A(t) = A_1(t) + A_2(t) + \cdots + A_k(t)$. Introducing the steady state probabilities

$$p_{1j_1j_2\ldots j_k}(x)dx = \lim_{t\to\infty} P(C(t) = 1,\ A_1(t) = j_1, \ldots,\ A_k(t) = j_k,\ x < z(t) \le x + dx),$$

$$p_{ij_1j_2\ldots j_k} = \lim_{t\to\infty} P(C(t) = i,\ A_1(t) = j_1,\ A_2(t) = j_2, \ldots, A_k(t) = j_k),$$

$$x \ge 0,\ i = 0, 1,\ \sum_{l=1}^{k} j_l \le N-1,\ j_l \ge 0, 1 \le l \le k,$$

for the joint stationary distribution p_{ij} of the server state and the orbit size (1) it holds

$$p_{ij} = \sum_{j_1+j_2+\cdots+j_k=j}^{N-1} p_{ij_1j_2\ldots j_k}.$$

The equations of statistical equilibrium are very complicated even for small k. For $k = 1$ this is the system (2) - (4) with $p = 0$, for $k = 2$ the system has the form

$$\frac{d}{dx}p_{1j_1j_2}(x) = -[(N-j-1)\lambda + \eta(x) + j\mu]p_{1j_1j_2}(x) + (N-j)\lambda p_{1,j_1-1,j_2}(x)$$

$$+(j_2+1)\mu p_{1,j_1,,j_2+1}(x) + (j_1+1)\mu p_{1,j_1+1,j_2-1}(x),$$

$$[j\mu + (N-j)\lambda]p_{0j_1j_2} = \int_0^\infty p_{1j_1j_2}(x)\eta(x)dx,$$

$$p_{1j_1j_2}(0) = (N-j)\lambda p_{0j} + (j_2+1)\mu p_{0,j_1,j_2+1} + (j_1+1)\mu p_{0,j_1+1,j_2},$$

$$j = j_1 + j_2,\ p_{1j_1j_2}(x) = p_{1j_1j_2} = 0 \text{ if } j_k \notin [0, N-1],\ k = 1, 2.$$

3. SIMULATIONS

In this section, on the basis of obtained via simulations results we study the properties of the orbit size and the server state. The simulations are performed in the case of exponentially distributed service times. Assuming that in the beginning of the process the server is free and there are no active subscribers, the simulation follows the short time changes of the system state as the direction of transformations is chosen according to the probability of each transformation and using generated random numbers. The method is described in detail in [2].

Five examples are presented as illustrations: simulation of the first restricted model with probability $p = 0.5$, of the second restricted model for $k = 1$, $k = 2$ and $k = 3$ and of model without restrictions. The observed mean $E[A]$, standard deviations $std[A]$ and the partition $E[A_1]$, $E[A_2]$, $E[A_3]$ (in case $k = 2$ and $k = 3$) of the orbit size as well as the probability P_1 of a busy channel are given in table 1. The corresponding theoretically obtained values in the system without losses (calculated according to the derived in [3] formulas) are also given. The values of the parameters are the same in each of these examples: $\lambda = 0.012$, $\mu = 0.1$, $\nu = 1$, $N = 50$, $n = 10000$, where n is the duration of the simulation (the number of time steps). The results presented in the table show that the type of the model greatly influences the observed characteristics, especially the mean number of the active subscribers.

Table1

	P_1	$E[A]$ (st.dev.)	$E[A_1]$	$E[A_2]$	$E[A_3]$
I restr. model, $p = 0.5$	0.69	9.22 (2.97)	-	-	-
II restr. model, $k = 1$	0.66	6.35 (2.5)	-	-	-
II restr. model, $k = 2$	0.70	10.11 (3.16)	6.06	4.05	-
II restr. model, $k = 3$	0.72	12.73 (3.16)	5.88	4.07	2.77
Unrestr. model	0.73	18.93 (4.02)	-	-	-
Theor. results (unrestr. m.)	0.72	19.10 (4.15)	-	-	-

In figures 1 - 2 we can see the effect not only of the system's type but also of the system's parameters on the considered characteristics. The graphs on fig.1 represent the dependence of the active subscribers mean percentage $E[A].100/N$ on the system's parameters: λ (the upper-left corner), μ (the upper-right corner), ν (the lower-left corner) and N (the lower-right corner). The dashed lines show the theoretically obtained values of $E[A].100/N$ in the unrestricted model (when active subscribers repeat their attempts until enter service) and the solid lines - the corresponding empirical means for the second restricted model with $k = 2$ (when active subscribers leave the orbit if their third attempt fails), each one obtained via simulations with 10000 steps. Fig.2 has the same structure as Fig.1, but it concerns the busy channel probability P_1. As we can see the behaviour of $E[A].100/N$ and P_1 in both considered models is similar, but the difference between these models is clearly shown as well.

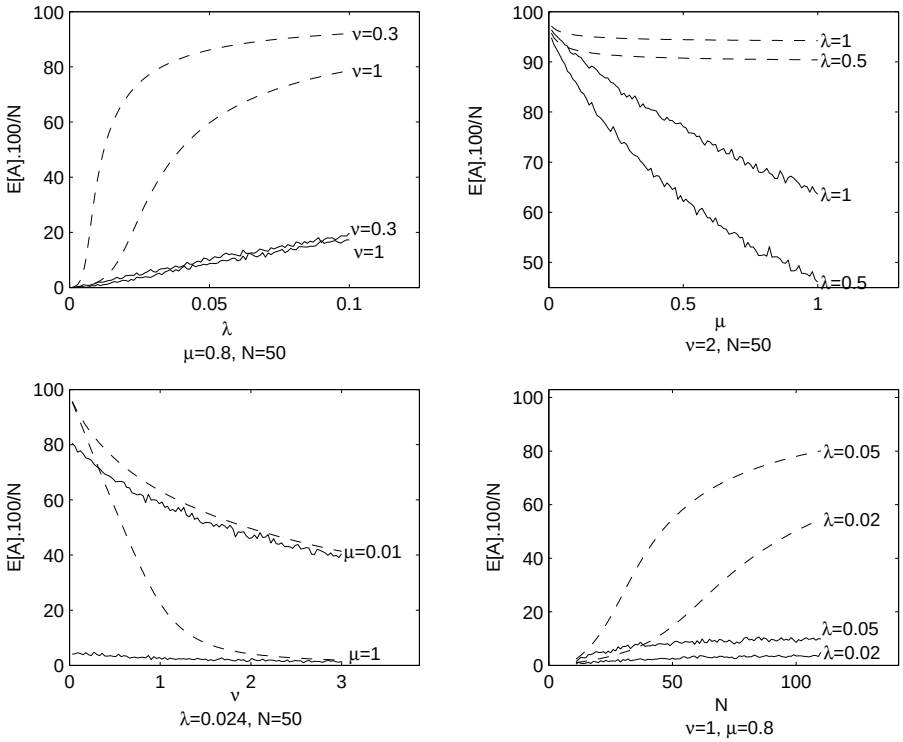
E[A].100/N
ν=0.3
ν=1
ν=0.3
ν=1
λ
μ=0.8, N=50
λ=1
λ=0.5
λ=1
λ=0.5
μ
ν=2, N=50
μ=0.01
μ=1
ν
λ=0.024, N=50
λ=0.05
λ=0.02
λ=0.05
λ=0.02
N
ν=1, μ=0.8

Fig. 1.

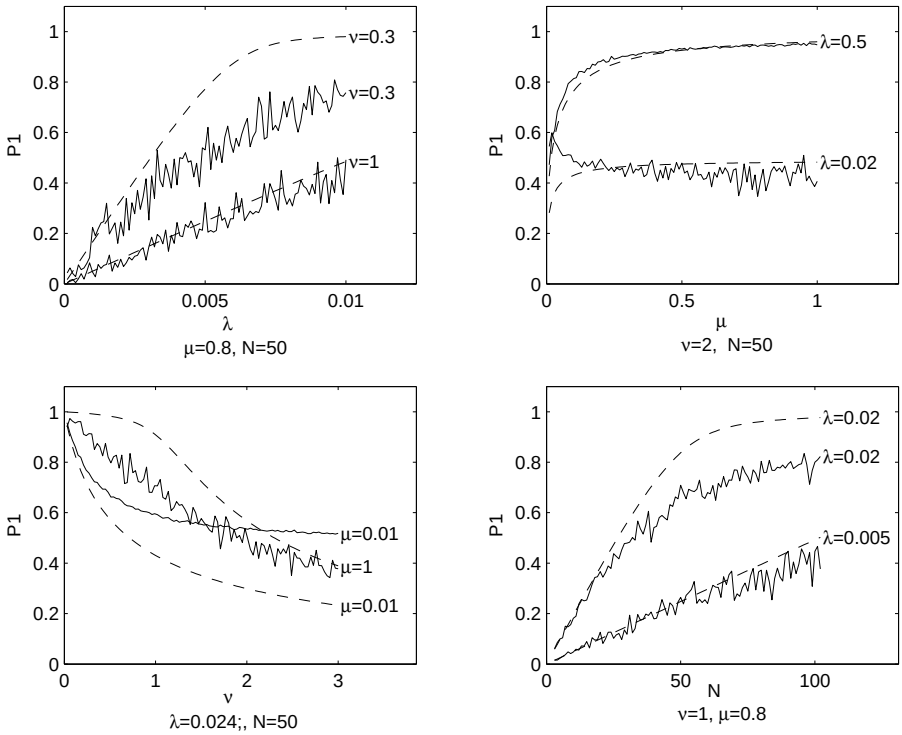
P1
ν=0.3
ν=0.3
ν=1
λ
μ=0.8, N=50
λ=0.5
λ=0.02
μ
ν=2, N=50
μ=0.01
μ=1
μ=0.01
ν
λ=0.024;, N=50
λ=0.02
λ=0.02
λ=0.005
N
ν=1, μ=0.8

Fig. 2.

REFERENCES

1. Artalejo J.R., Gómez-Corral A. Retrial Queueing Systems: A Computational Approach. Berlin Heidelberg: Springer, 2008.
2. Dragieva V.I. Queue length simulations in a finite single-line queueing system with repeated calls // Pliska Studia Mathematica Bulgarica. 2009. 19. P. 121-134.
3. Dragieva V.I. A single-server queueing system with a finite source and repeated calls (Russian)// Problemy Peredachi Informacii. 1994. 30. no3. P. 104-111. Translation in Problems Inform. Transmission. 1995. 30. no3. P. 283-289.
4. Falin G.I., Artalejo J.R. A finite source retrial queue // European Journal of Operational Research. 1998. 108. P. 409-424.
5. Falin G.I., Templeton J.G.C. Retrial Queues. Chapman and Hall, 1997.
6. Janssens G.K. The quasi-random input queueing system with repeated attempts as a model for collision-avoidance star local area network // IEEE Transactions on Communications. 1997. 45. P. 360-364.
7. Kornyshev Y.N. Design of a fully accessible switching system with repeated calls // Telecommunications. 1969. 23. P. 46-52.
8. Mehmet-Ali M.K., Elhakeem A.K., Hayes J.F. Traffic analysis of a local area network with star topology // IEEE Transactions on Communications. 1988. 36. P. 703-712.
9. Ohmura H., Takahashi Y. An analysis of repeated call model with a finite number of sources // Electronics and Communications in Japan. 1985. 68. P. 112-121.

"SIRIUS-T" PACKAGE FOR CHARACTERISTICS EVALUATION OF THE QUEUEING SYSTEMS

A. Dudin[1], V. Klimenok[1], G. Tsarenkov[2,⋆]

[1] *Belarusian State University,*

[2] *BestSoft FLCC*

[1,2] *Minsk, Belarus*

[⋆] gtsarenkov@tut.by

This paper contains description of new features of software package "Sirius-T" as an elaboration of "Sirius++" and "Sirius-C" packages. "Sirius-T" introduces new models for calculation characteristics of tandem queues allowing performance evaluation and capacity planning of the telecommunication networks and their fragments. Additionally, the software facilitates a study of queueing models under consideration of correlated service and evaluation of the algorithms' numerical complexity.

Keywords: Telecommunications, Computer Networks, Performance Characteristics, Queueing Systems, Tandems.

1. INTRODUCTION

The described software "Sirius-T" is the next generation of the software "Sirius++" that is described in [5]. Apart minor extensions that facilitate study of the queueing systems under consideration of the correlated service (described in terms of semi-Markovian process) some major improvements have been made.

One of such improvements allows analysis of numerical complexity of algorithms. Due to approximate nature of the implemented algorithms there is no analytical way known to perform such analysis. However, with minor performance loss it is possible to gather the information during calculation.

Another improvement allows using calculation under different platforms (e.g. different operating systems or hardware). This is important while calculation precision differs depending on the development tools shipped with the platform.

Finally, the most significant improvement was achieved by adding support of new queueing systems. In particular several tandem queues have been added to a set of supported models. This allows studying of more complex networks and their fragments.

The software "Sirius-T" has been developed in C++ using Microsoft Visual Studio 2008 and GNU GCC 4.5.1.

2. DESCRIPTION OF THE TANDEM QUEUES

Before describing tandem models supported by the "Sirius-T", let us briefly specify common aspects of studied tandem queues. The key difference of tandems from single

server queues is that it has two service devices (servers). Service time duration at the first server has general distribution with the first moment b_1, $0 < b_1 < \infty$. The first server has a buffer of size N (two cases are considered $N < \infty$ and $N = \infty$).

The second server is characterized by the PH-type service time distribution having an irreducible representation (β, S). Here β is the stochastic row-vector of dimension K and S is $K \times K$ matrix having the negative diagonal and non-negative non-diagonal entries, such as the column vector $S_0 = -S\mathbf{e}$ is non-negative and has at least one positive entry. The average service time is defined as $\beta(-S)^{-1}\mathbf{e}$. Here and in the sequel $\mathbf{e}$ is column vector of appropriate size consisting of units. For more information about PH see [7]. The buffer before the second server has size $M-1$, $M < \infty$.

Each request needs to be sequentially served by first and second servers.

"Sirius-T" continues utilizing batch Markovian Arrival process ($BMAP$) as an input flow. The $BMAP$, a special class of tractable Markov renewal process, is a rich class of point processes. It includes many well-known processes such as stationary Poisson, PH-renewal process, Markov Modulated Poisson Process ($MMPP$) and others. The epochs of customers' arrival (possibly in batches of a random size) coincide with the transition epochs of some continuous-time Markov chain $\nu_t, t \geq 0$ called as the directing process of the BMAP. This process has a finite state space $\{0, \ldots, W\}$ and it is completely defined by some set of $(W+1) \times (W+1)$ matrices $D_k, k \geq 0$. The matrix D_k consists of intensities of transitions of the process $\nu_t, t \geq 0$ that are accompanied by arrival of a batch of size k into the system. The BMAP was introduced by D.Lucantoni in [6] as a more nice form of versatile input process introduced earlier by M.Neuts in [8]. The BMAP is recommended by many researchers as a good descriptor of flows in the modern telecommunication networks. It takes well into account the bursty, correlated nature of these flows. It makes its exploiting to model the real life flows being attractive.

For understanding the examples from the following section we need to describe some additional characteristics of the BMAP flow. A matrix generating function

$$D(z) = \sum_{k=0}^{\infty} D_k z^k, \qquad |z| \leq 1,$$

introduced by D.Lucantoni in [6] facilitates analytical study of queueing systems. The vector $\boldsymbol{\theta}$ of the process ν_t, $t > 0$ stationary distribution satisfies equations $\boldsymbol{\theta} D(1) = \mathbf{0}, \boldsymbol{\theta}\mathbf{e} = 1$. Here $\mathbf{0}$ is zero row vector of appropriate size. The average intensity λ (fundamental rate) of the $BMAP$ is defined as

$$\lambda = \boldsymbol{\theta} D'(z)|_{z=1}\mathbf{e}.$$

The intensity λ_g of groups arrival is defined as

$$\lambda_g = \boldsymbol{\theta}(-D_0)\mathbf{e},$$

variance c_{var} of intervals between the groups arrival is calculated as:

$$c_{var} = 2{\lambda_g}^{-1}\boldsymbol{\theta}(-D_0)^{-1}\mathbf{e} - {\lambda_g}^{-2},$$

the correlation coefficient c_{cor} of intervals between the successive groups arrival is calculated as

$$c_{cor} = ({\lambda_g}^{-1}\boldsymbol{\theta}(-D_0)(D(1) - D_0)(-D_0)^{-1}\mathbf{e} - {\lambda_g}^{-2})/v.$$

We consider partial admission losses at first server if $N < \infty$, i.e. the only requests from batch arrival will be lost that exceed the free buffer capacities.

We can now describe and illustrate with numerical examples models supported by "Sirius-T".

3. TANDEM QUEUES WITH LOSSES

We consider $BMAP|G|1|N \to \cdot|PH|1|M$ tandem with possible losses at the first and the second server (please see the figure 1 for illustration). This model was studied in paper [2].

In case the entering batch of customers finds insufficient number of places in a buffer (or the buffer is already full at all), the appropriate number of customers from the batch joins a queue while the rest (or even the whole group) leaves the system forever (i.e. is lost at the first server). We denote the probability $P_{loss}^{(1)}$ of the request to be lost at the first server. In case the customer completes the service at first server and meets the buffer before the second server be busy, this customer leaves the system forever and is considered to be lost at the second server. We denote the probability $P_{loss}^{(2)}$ of the request to be lost at the second server. If buffer before first server is infinite, then requirements can be lost at the second server only.

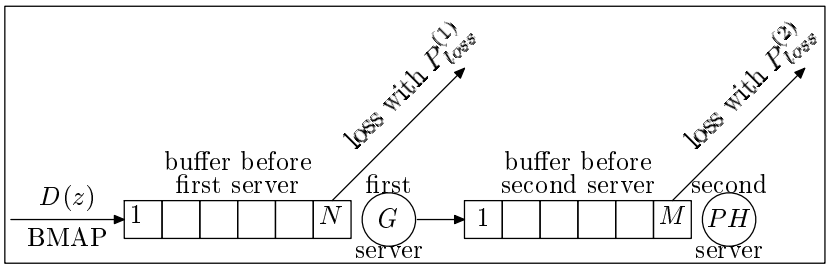

Fig. 1. Structure of the tandem queue with losses

Let us illustrate the module with some example.

The input parameters of the tandem are defined as specified below.

Table 1 $BMAP$**-flow of intensity** $\lambda = 10$**, intensity of groups** $\lambda_g = 5$**, with correlation** $c_{cor} = 0.2$**, and variation** $c_{var} = 12.2732$

D_0		$D_1 = D_3$		D_2	
-6.74538	5.45412×10^{-6}	2.01021	0.0134084	2.68027	0.0178778
5.45412×10^{-6}	-0.219455	0.036728	0.0291068	0.0489707	0.038809

Service time at the first server is degenerate. To vary service time the following values are used $\{0.01, 0.03, 0.05, 0.07, 0.08, 0.085, 0.09, 0.095, 0.1, 0.12, 0.14\}$.

Phase-type service time distribution at the second server has the following parameters

$$S = \begin{bmatrix} -20 & 0 \\ 0 & -80 \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix}.$$

These parameters make mean service time at the second phase equal to 0.03875.

The buffer capacity M at the second server is 2.

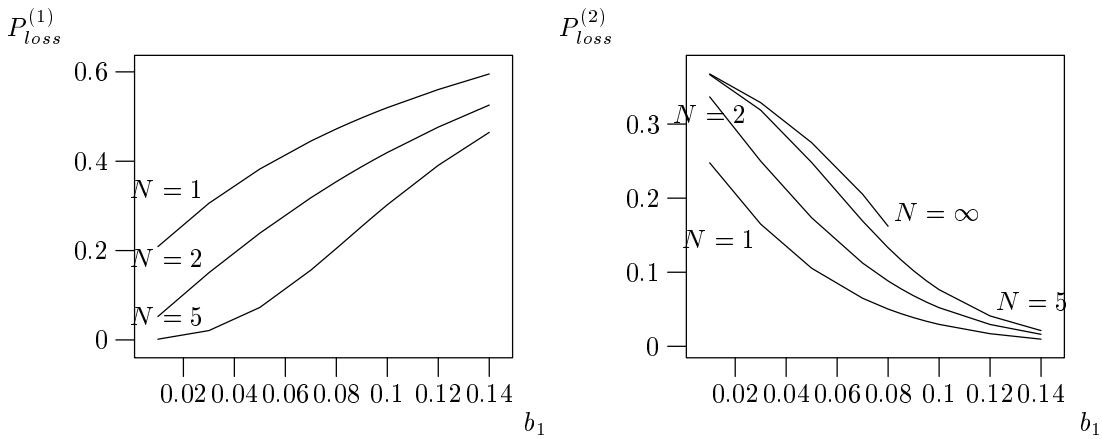

Fig. 2. Loss probability $P_{loss}^{(1)}$ and $P_{loss}^{(2)}$ at the first and at the second servers depending on average service time b_1 at the first phase

4. TANDEM QUEUES WITH BLOCKINGS

We consider $BMAP|G|1|N \to \cdot|PH|1|M$ tandem with possible losses at the first server and blockings of the first server if there is no free space in a buffer at the second server (please see the figure 3 for illustration). This model was studied in paper [1].

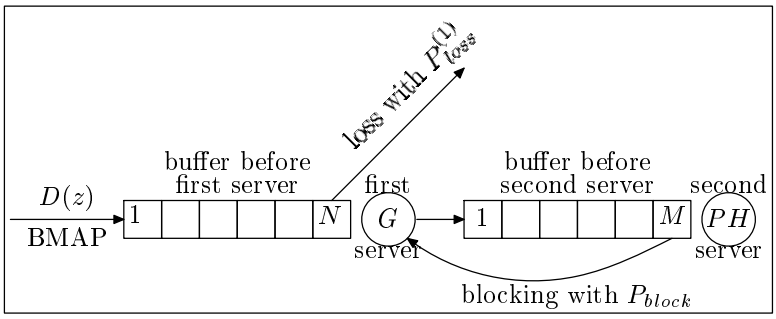

Fig. 3. Structure of the tandem queue with blockings

Let us illustrate the module with some example. We will use input data from the previous section. However, due to ergodicity condition, in case of an inifinite first buffer, we need to limit service time at the first server to the following values $\{0.01, 0.03, 0.05, 0.07, 0.08\}$.

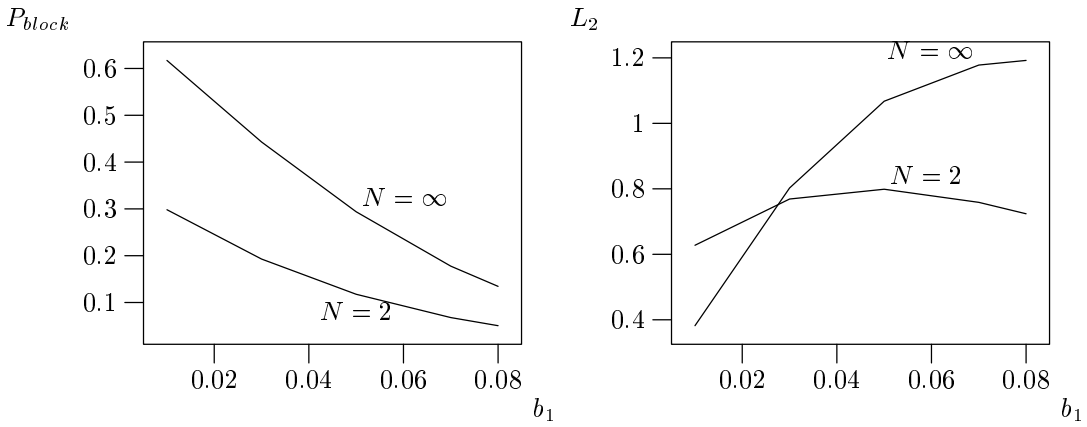

Fig. 4. Blocking probability P_{block} and average queue length L_2 at the second server depending on average service time b_1 at the first phase

5. TANDEM QUEUES WITH FEEDBACK AND LOSSES

We consider $BMAP|G|1|N \to \cdot|PH|1|M$ tandem with possible losses at the first server and at the second server as well as feedback mechanism (please see the figure 3 for illustration). By feedback we meen, that any request with probability p after been served at the second server can return to the buffer at the first server, and with probability $1-p$ can leave system forever. This model was studied in paper [4].

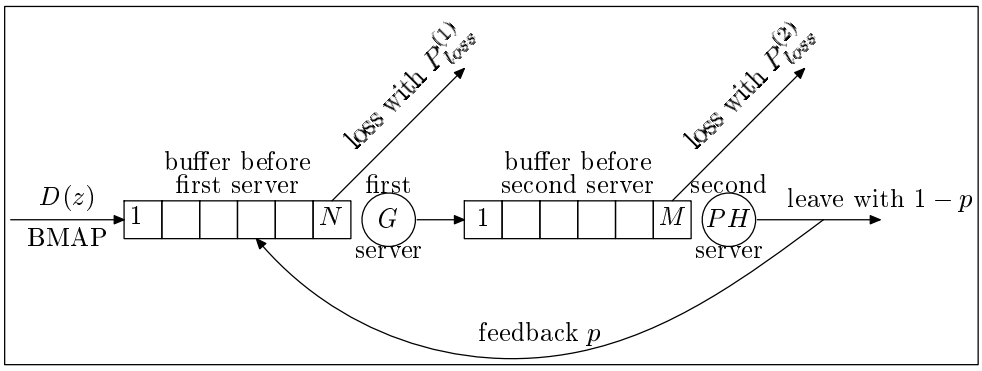

Fig. 5. Structure of the tandem queue with feedback and losses

Let us illustrate the module with some example. We will use input parameters as in the second example, but the service time distribution at the first server we consider deterministic with $T = 0.05$.

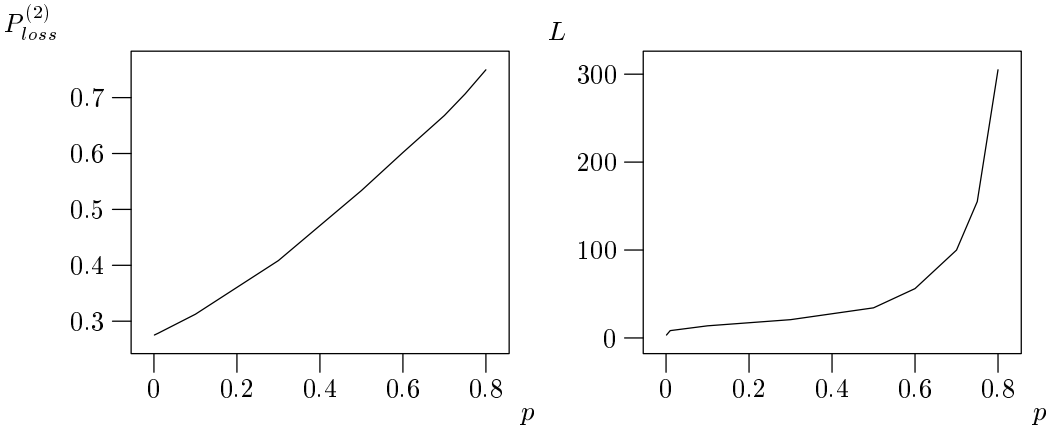

Fig. 6. Loss probability $P_{loss}^{(2)}$ and average queue length L_2 at the second server depending on feedback probability p

6. TANDEM QUEUES WITH RETRIALS AND LOSSES

We consider $BMAP|G|1 \to \cdot|PH|1|M$ tandem with retrials at the first server and possible losses at the second server. The orbit before the first server is infinite. If an arriving request finds the first server busy it joins the orbit and tries to get service in exponentially distributed time intervals with intensity $\alpha_i = i\alpha + \gamma$, where i denotes number of customers at the orbit. This model was studied in paper [3].

We will use the same input data as in previous section, but the retrial intensity in this example is considered to be linear one, i.e.

$$\alpha_i = i\alpha + \gamma, \qquad i \geq 1,$$

where i denotes number of customers in an orbit and $\alpha = 3$, $\gamma = 5$.

The buffer size at the second server as well as the calculated performance characteristics are given in the table 2 below.

Additionally, probability $P_0^{(2)}$ denotes that the second server is idle at arbitrary epoch.

Table 2 **Results of some performance characteristics calculations for different input parameters**

Parameters	Case 1	Case 2
M	9	4
L_1	1.96517	1.13647
L_2	2.33711	1.098326
P_{loss}	0.178394	0.254642
$P_{0,0}$	0.279899	0.286573
$P_0^{(2)}$	0.308468	0.281546

7. CONCLUSION

The new "Sirius-T" software package allows study of new models to calculate performance characteristics of the tandem queueing systems as a adequate models of the fragments of the telecommunication networks. Apart that, it extends algorithms with additional characteristics allowing investigation of numerical complexity. All this makes "Sirius-T" application valuable not just for sceintists, but also for engineers who needs to perform analysis of some telecommunication networks or their fragments.

The development of the "Sirius-T" will be continued to allow handling more complex systems and algorithms.

REFERENCES

1. *Breuer L., Dudin A.N., Kliminok V.I., Tsarenkov G.V.* A two-phase $BMAP|G|1|N \to PH|1|M-1$ with blocking // Automation and Remote Control. 2004. № 1. P. 117-130. (in Russian).
2. *Dudin A.N., Breuer L., Klimenok V.I., Tsarenkov G.V.* Tandem queue $BMAP|G|1|\overline{N} \to \cdot|PH|1|M$ with losses // Performance Evaluation. 2005. № 61. P. 17-40.
3. *Dudin A.N., Klimenok V.I., Che Soong Kim, Tsarenkov G.V.* Investigation of the $BMAP/G/1 \to \cdot/PH/1/M-1$ tandem queue with retrials and losses // Applied Mathematical Modelling, 2010. V. 34, Issue 10. P. 2926-2940.
4. *Dudin A.N., Klimenok V.I., Che Soong Kim, Tsarenkov G.V.* The $BMAP/G/1 \to \cdot/PH/1/M-1$ tandem queue with feedback and losses // Performance Evaluation. 2007. V. 64, Issue 7-8. P. 802-818.
5. *Dudin A. N., Klimenok V. I.,, Tsarenkov G. V.* Software "Sirius++" for Performance Evaluation of Modern Communication Networks // Modeling and Simulation. 16th European Simulation Multiconference, 2002. P. 489-493.
6. *Lucantoni D.* New results on the single server queue with a batch markovian arrival process // Communications in Statistics - Stochastic Models, 1991. V. 7. № 1. P. 1-46.
7. *Neuts M. F.* Matrix-geometric solutions in stochastic models // Baltimore, The Johns Hopkins University Press. 1981.
8. *Neuts M. F.* Structured stochastic matrices of M/G/1 type and their applications // New York, M. Dekker. 1989.

СИСТЕМА ОБСЛУЖИВАНИЯ $MAP|PH|N|R-N$ КАК МОДЕЛЬ ФУНКЦИОНИРОВАНИЯ КОНТАКТ-ЦЕНТРА

С. Дудин, О. Тарамин*

Белорусский государственный университет
Минск, Беларусь
* taramin@mail.ru

В статье исследуется многолинейная система массового обслуживания с конечным буфером и нетерпеливыми запросами. На вход системы поступает марковский поток запросов. Времена обслуживания запросов каждым прибором имеют распределения фазового типа. Во время ожидания обслуживания в буфере запросы могут проявлять нетерпеливость и покидать систему.

Данная работа была частично поддержана Белорусским республиканским фондом фундаментальных исследований через грант № Ф10МЛД-010.

Ключевые слова: марковский поток, процесс обслуживания фазового типа, нетерпеливые запросы, контакт-центр.

1. ВВЕДЕНИЕ

В данной статье исследуется многолинейная система обслуживания с нетерпеливыми запросами. Эта система может использоваться для моделирования и оптимизации функционирования контакт-центра. Обслуживание абонентов осуществляется несколькими операторами. В случае занятости всех операторов абоненту предлагается подождать, при этом ему сообщается его номер в очереди и ориентировочное время ожидания. В зависимости от его номера в очереди абонент принимает решение, ожидать или нет обслуживания. Во время ожидания абонент также может разрывать соединение из-за нетерпеливости.

Считаем, что от обслуживания каждого абонента система получает прибыль, в то время как содержание оператора требует затрат. Необходимо определить такое число операторов, при котором прибыль центра была бы максимальной.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается система обслуживания, состоящая из N приборов и буфера размера $R-N$, $R \geq N$. В систему поступает марковский входной поток (MAP – Markovian Arrival Process) запросов, заданный неприводимой цепью Маркова ν_t, $t \geq 0$, с непрерывным временем и конечным пространством состояний $\{0, ..., W\}$.

Время пребывания цепи в состоянии ν экспоненциально распределено с положительным параметром λ_ν. Когда время пребывания в состоянии ν истекло, с вероятностью $p^{(k)}_{\nu,\nu'}$ процесс ν_t переходит в состояние ν' и при этом генерируется группа из k запросов, $k = 0, 1, \nu, \nu' = \overline{0, W}$. Поведение MAP-потока полностью характеризуется матрицами $D_k, k = 0, 1$, которые определяются следующим образом

$$(D_k)_{\nu,\nu'} = \lambda_\nu p^{(k)}_{\nu,\nu'}, k = 1 \text{ и } k = 0, \nu \neq \nu', \ (D_0)_{\nu,\nu} = -\lambda_\nu, \nu = \overline{0, W}.$$

Матрица $D(1) = D_0 + D_1$ представляет собой инфинитезимальный генератор цепи $\nu_t, t \geq 0$.

Средняя интенсивность поступления запросов λ имеет вид

$$\lambda = \boldsymbol{\chi} D_1 \mathbf{e},$$

где $\boldsymbol{\chi}$ - вектор стационарного распределения цепи Маркова $\nu_t, t \geq 0$. Вектор $\boldsymbol{\chi}$ является единственным решением системы линейных алгебраических уравнений $\boldsymbol{\chi} D(1) = \mathbf{0}$, $\boldsymbol{\chi}\mathbf{e} = 1$. Здесь и далее $\mathbf{e}$ – вектор-столбец, состоящий из единиц, $\mathbf{0}$ – вектор-строка, состоящая из нулей.

Коэффициент вариации c_{var} длин интервалов между моментами поступления запросов

$$c^2_{var} = 2\lambda \boldsymbol{\chi}(-D_0)^{-1}\mathbf{e} - 1.$$

Коэффициент корреляции c_{cor} длин двух соседних интервалов между поступлением запросов

$$c_{cor} = (\lambda \boldsymbol{\chi}(-D_0)^{-1}(D(1) - D_0)(-D_0)^{-1}\mathbf{e} - 1)/c^2_{var}.$$

Время обслуживания запроса каждым прибором имеет распределение фазового. Время обслуживания, имеющее распределение фазового типа (PH – Phase-type distribution) с неприводимым представлением $(\boldsymbol{\beta}, S)$, можно интерпретировать как время, за которое цепь Маркова с непрерывным временем $\eta_t, \ t \geq 0$, имеющая несущественные состояния $\{1, \ldots, M\}$ и поглощающее состояние $M + 1$, достигнет поглощающего состояния. Начальное состояние цепи Маркова $\eta_t, \ t \geq 0$, в момент начала обслуживания запроса определяется вероятностным вектором $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$. Переходы цепи $\eta_t, \ t \geq 0$, которые не приводят к окончанию обслуживания, задаются субгенератором S размера $M \times M$. Интенсивности переходов в поглощающее состояние описываются вектором $\mathbf{S}_0 = -S\mathbf{e}$. Среднее время обслуживания определяется формулой $b_1 = \boldsymbol{\beta}(-S)^{-1}\mathbf{e}$.

Если все приборы заняты и число запросов в буфере в момент прихода запроса равно i, $i = \overline{0, R - N - 1}$, то этот запрос с вероятностью q_i покидает систему, а с дополнительной вероятностью становится в очередь.

Если в момент прихода запроса свободные места в буфере отсутствуют, запрос покидает систему. Через экспоненциально распределенное с параметром $\alpha, 0 < \alpha < \infty$, время с момента попадания в буфер, запрос покидает систему, если он до этого не попал на обслуживание.

Качество функционирования системы описывается критерием

$$J(N) = a\lambda_{out} - b\lambda P^{(loss)} - dN,$$

который определяет прибыль системы в единицу времени. Здесь a – прибыль, полученная от обслуживания одного запроса, λ_{out} – среднее число запросов, обслуженных системой в единицу времени, b – штраф за потерю запроса, $P^{(loss)}$ – вероятность потери произвольного запроса, d – стоимость содержания одного прибора в единицу времени.

Необходимо определить такое число приборов N, при котором достигает максимума критерий $J(N)$.

3. ПРОЦЕСС ИЗМЕНЕНИЯ СОСТОЯНИЙ СИСТЕМЫ И СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ЧИСЛА ЗАПРОСОВ В СИСТЕМЕ

Пусть i_t, $i_t \in \{0, \ldots, R\}$, – число запросов в системе,

$\nu_t \in \{0, \ldots, W\}$, – состояние управляющего процесса MAP-потока поступления запросов,

$\eta_t^{(l)}$, $\eta_t^{(l)} \in \{1, \ldots, M\}$, – состояние управляющего процесса обслуживания l-м занятым прибором в момент времени t, $t \geq 0$, $l \in \{1, \ldots, \min\{i_t, N\}\}$.

Процесс $\xi_t = \{i_t, \nu_t, \eta_t^{(1)}, \ldots, \eta_t^{(\min\{i_t, N\})}\}$, $t \geq 0$, является регулярной неприводимой цепью Маркова с непрерывным временем.

Занятые приборы предполагаются перенумерованными в порядке их занятия: номер 1 имеет наиболее долготекущий к данному моменту времени процесс обслуживания. Соответственно максимальный номер имеет процесс обслуживания, которые стартовал позже всех остальных. При завершении процесса обслуживания запроса и отсутствии очереди соответствующий ему номер прибора вычеркивается из записи, и производится перенумерация.

Перенумеруем состояния цепи ξ_t в лексикографическом порядке. Множество состояний, имеющих значение (i, ν) двух первых компонент цепи, будем называть макросостоянием (i, ν).

Пусть Q – генератор цепи Маркова ξ_t, $t \geq 0$, сформированный из блоков $Q_{i,j}$, состоящих из матриц $(Q_{i,j})_{\nu,\nu'}$ интенсивностей переходов цепи ξ_t, $t \geq 0$, из макросостояния (i, ν) в макросостояние (j, ν'), $\nu, \nu' = \overline{0, W}$.

Введем следующие обозначения:

- $A^{\oplus\, i} = \underbrace{A \oplus A \oplus \ldots \oplus A}_{i}, i > 0$, $A^{\oplus\, 0} = O_{1\times 1}$;
- $\bar{W} = W + 1$;
- $\alpha_i = (i - N)\alpha$.

Лемма 1. *Генератор Q имеет следующую блочно-трехдиагональную структуру:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & \dots & O & O \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \dots & O & O \\ O & Q_{2,1} & Q_{2,2} & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & Q_{R-1,R-1} & Q_{R-1,R} \\ O & O & O & \dots & Q_{R,R-1} & Q_{R,R} \end{pmatrix}.$$

Ненулевые блоки генератора имеют следующий вид:

$$Q_{i,i} = D_0 \oplus S^{\oplus i},\ i = \overline{0, N-1},$$

$$Q_{N,N} = D_0 \oplus S^{\oplus N} + q_0 D_1 \otimes I_{M^N},$$

$$Q_{i,i} = D_0 \oplus S^{\oplus N} + q_{i-N} D_1 \otimes I_{M^N} - \alpha_i I_{\bar{W}M^N},\ i = \overline{N+1, R-1},$$

$$Q_{R,R} = D(1) \oplus S^{\oplus N} - \alpha_i I_{\bar{W}M^N},$$

$$Q_{i,i-1} = I_{\bar{W}} \otimes \mathbf{S_0}^{\oplus i},\ i = \overline{1, N},$$

$$Q_{i,i-1} = I_{\bar{W}} \otimes (\mathbf{S_0}\boldsymbol{\beta})^{\oplus N} + \alpha_i I_{\bar{W}M^N},\ i = \overline{N+1, R},$$

$$Q_{i,i+1} = D_1 \otimes I_{M^i} \otimes \boldsymbol{\beta},\ i = \overline{0, N-1},$$

$$Q_{i,i+1} = (1 - q_{i-N}) D_1 \otimes I_{M^N},\ i = \overline{N, R-1}.$$

Доказательство. Доказательство леммы опирается на анализ переходов цепи Маркова ξ_t, $t \geq 0$, за бесконечно малый интервал времени с последующей группировкой интенсивностей соответствующих переходов в блоки матрицы Q. □

Так как цепь Маркова $\xi_t = \{i_t, \nu_t, \eta_t^{(1)}, \dots, \eta_t^{(\min\{i_t,N\})}\}$, $t \geq 0$, регулярная и неприводимая, а также имеет конечное пространство состояний, то существуют следующие пределы (стационарные вероятности):

$$\pi(i, \nu, \eta^{(1)}, \dots, \eta^{(\min\{i,N\})}) = \lim_{t\to\infty} P\{i_t = i,\ \nu_t = \nu,\ \eta_t^{(1)} = \eta^{(1)}, \dots, \eta_t^{(\min\{i_t,N\})} = \eta^{(\min\{i,N\})}\},$$

$$i = \overline{0, R},\ \nu = \overline{0, W},\ \eta^{(l)} = \overline{1, M},\ l \in \{1, \dots, \min\{i, N\}\}.$$

Перенумеруем стационарные вероятности цепи ξ_t в лексикографическом порядке и сформируем из них векторы $\boldsymbol{\pi}(i, \nu)$, состоящие из вероятностей $\pi(i, \nu, \eta^{(1)}, \dots, \eta^{(\min\{i,N\})})$, записанных в лексикографическом порядке по компонентам $(\eta^{(1)}, \dots, \eta^{(\min\{i,N\})})$, и векторы

$$\boldsymbol{\pi}_i = (\boldsymbol{\pi}(i, 0), \boldsymbol{\pi}(i, 1), \dots, \boldsymbol{\pi}(i, W)),\ i = \overline{0, R}.$$

Известно, что вектор-строка $(\boldsymbol{\pi}_0, \dots, \boldsymbol{\pi}_R)$ – единственное решение системы

$$(\boldsymbol{\pi}_0, \dots, \boldsymbol{\pi}_R) Q = \mathbf{0},\ \ (\boldsymbol{\pi}_0, \dots, \boldsymbol{\pi}_R)\mathbf{e} = 1. \tag{1}$$

В случае, когда размерность системы (1) невелика, эту систему легко решить с помощью компьютера стандартными методами. В противном случае, может быть

использован специальный устойчивый алгоритм, решающий систему такого вида, который был разработан в [1].

С учетом того, что генератор Q в данной модели имеет блочно-трехдиагональную форму, в то время как в [1] допускается блочная верхне-Хессенбергова форма, алгоритм из [1] можно записать в более простом виде.

Теорема 1. *Компоненты $\boldsymbol{\pi}_i$, $i = \overline{0, R}$, вектора решения системы (1) связаны соотношениями:*

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_0 F_i, i = \overline{1, R},$$

где матрицы F_i вычисляются рекуррентно:

$$F_0 = I, \;\; F_i = -F_{i-1} Q_{i-1,i} (Q_{i,i} + Q_{i,i+1} G_i)^{-1}, i = \overline{1, R},$$

матрицы $G_i, i = \overline{0, R-1}$, вычисляются с помощью обратной рекурсии:

$$G_i = -(Q_{i+1,i+1} + Q_{i+1,i+2} G_{i+1})^{-1} Q_{i+1,i}, \;\; i = R-2, R-3, \ldots, 0,$$

при начальном условии

$$G_{R-1} = -(Q_{R,R})^{-1} Q_{R,R-1},$$

а вектор $\boldsymbol{\pi}_0$ является единственным решением следующей системы:

$$\boldsymbol{\pi}_0 (Q_{0,0} + Q_{0,1} G_0) = \mathbf{0}, \quad \boldsymbol{\pi}_0 \sum_{l=0}^{R} F_l \mathbf{e} = 1.$$

Доказательство. Доказательство вытекает непосредственно из результатов, полученных в [1]. □

4. ХАРАКТЕРИСТИКИ ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМЫ

Найдя векторы стационарных вероятностей $\boldsymbol{\pi}_i$, $i = \overline{0, R}$, можно вычислить различные характеристики производительности системы.

Распределение вероятностей числа запросов в системе вычисляется как

$$\lim_{t \to \infty} P\{i_t = i\} = \boldsymbol{\pi}_i \mathbf{e}, \; i = \overline{0, R}.$$

Среднее число занятых приборов

$$\bar{N} = \sum_{i=1}^{N} i \boldsymbol{\pi}_i \mathbf{e} + N \sum_{i=N+1}^{R} \boldsymbol{\pi}_i \mathbf{e}.$$

Интенсивность выходящего потока обслуженных запросов

$$\lambda_{out} = \sum_{i=0}^{N} \boldsymbol{\pi}_i (\mathbf{e}_{\bar{W}} \otimes \sum_{l=0}^{i-1} (\mathbf{e}_{M^l} \otimes \mathbf{S}_0 \otimes \mathbf{e}_{M^{i-l-1}})) + \sum_{i=N+1}^{R} \boldsymbol{\pi}_i (\mathbf{e}_{\bar{W}} \otimes \sum_{l=0}^{N-1} (\mathbf{e}_{M^l} \otimes \mathbf{S}_0 \otimes \mathbf{e}_{M^{N-l-1}})).$$

Вероятность потери запроса на входе в систему из-за заполненности буфера определяется как

$$P^{(ent-loss)} = \lambda^{-1}\boldsymbol{\pi}_R(D_1 \otimes I_{M^N})\mathbf{e}.$$

Вероятность ухода запроса на входе в систему из-за нежелания ожидать:

$$P^{(esc-loss)} = \lambda^{-1}\sum_{i=N}^{R-1} q_{i-N}\boldsymbol{\pi}_i(D_1 \otimes I_{M^N})\mathbf{e}.$$

Вероятность потери запроса:

$$P^{(loss)} = 1 - \frac{\lambda_{out}}{\lambda}.$$

Вероятность ухода запроса из очереди из-за нетерпеливости:

$$P^{(imp-loss)} = P^{(loss)} - P^{(ent-loss)} - P^{(esc-loss)}.$$

5. ЗАКЛЮЧЕНИЕ

В данной статье исследована многолинейная система массового обслуживания с конечным буфером и нетерпеливыми запросами. Построен процесс функционирования данной системы, найдено стационарное распределение вероятностей состояний системы, получены формулы для нахождения основных характеристик производительности.

ЛИТЕРАТУРА

1. *Klimenok V., Kim C.S., Orlovsky D., Dudin A.* Lack of invariant property of Erlang loss model in case of the MAP input // Queueing Systems. 2005. V. 49. P. 187–213.

ИНВАРИАНТНОСТЬ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ СОСТОЯНИЙ СЕТЕЙ С МНОГОРЕЖИМНЫМИ СТРАТЕГИЯМИ

А. Ерёмина

Гродненский государственный университет имени Янки Купалы

г. Гродно, Беларусь

aleksandrae@mail.ru

Рассматривается открытая сеть массового обслуживания, в которой циркулируют заявки различных типов. Количество работы по обслуживанию поступающего требования является случайной величиной с произвольной функцией распределения. Однолинейные узлы могут работать в нескольких режимах. Количество работы по переключению режима прибора также является случайной величиной с произвольной функцией распределения. Процессы обслуживания заявок и переключения режимов функционирования приборов не зависят друг от друга. Устанавливается инвариантность стационарного распределения состояний сети по отношению к функциональным формам распределений величин работ по обслуживанию заявок и переключению режимов при фиксированных первых моментах.

Ключевые слова: сеть массового обслуживания, многорежимное обслуживание, стационарное распределение, инвариантность.

1. ВВЕДЕНИЕ

Большую важность для практики представляет изучение сетей массового обслуживания, в которых обслуживающие приборы в узлах могут работать с различной интенсивностью, так как любые технические средства в процессе их эксплуатации могут полностью или частично выходить из строя, требовать замены или ремонта. В работе [1] были введены в рассмотрение сети массового обслуживания с многорежимными стратегиями, в узлах которых приборы могут функционировать в различных режимах. Каждый режим обслуживания характеризуется своими показателями, то есть при переходе прибора в более “худший” режим его производительность уменьшается. В [2] для открытой сети с многорежимными стратегиями, несколькими типами заявок и дисциплиной обслуживания LCFS PR было найдено стационарное распределение в мультипликативной форме. Однако считалось, что время обслуживания и время пребывания прибора в каждом режиме имеют экспоненциальное распределение. На практике указанные предположения чаще всего

не выполняются. Поэтому в [3] были рассмотрены аналогичные открытые сети, в предположении, что время пребывания в каждом режиме имеет показательное распределение, а количество работы по обслуживанию поступающих в узел заявок – произвольное. Было установлено, что стационарное распределение вероятностей состояний указанных сетей не зависит от вида законов распределения величин работ по обслуживанию заявок в узлах, если фиксированы первые моменты этих законов.

В настоящей статье исследуются открытые сети массового обслуживания, для которых количество работы по обслуживанию поступающего требования и количество работы по переключению с одного режима функционирования прибора на другой одновременно являются случайными величинами с произвольными функциями распределения. Устанавливается инвариантность стационарного распределения вероятностей состояний указанных сетей по отношению к виду распределений длительностей обслуживания и времен пребывания в режимах при фиксированных математических ожиданиях.

2. ПОСТАНОВКА ЗАДАЧИ

Рассматривается открытая сеть массового обслуживания, состоящая из N однолинейных узлов, в которой циркулируют заявки M типов. Поступающий в сеть поток заявок – простейший с параметром λ. Каждая заявка входного потока независимо от других заявок с вероятностью $p_{0(l,u)}$ направляется в l-й узел и становится заявкой u-го типа ($\sum\limits_{l=1}^{N}\sum\limits_{u=1}^{M} p_{0(l,u)} = 1,\ l = \overline{1,N},\ u = \overline{1,M}$). Заявка u-го типа, обслуженная в l-м узле, независимо от других заявок мгновенно с вероятностью $p_{(l,u)(k,v)}$ направляется в k-й узел и становится заявкой типа v, а с вероятностью $p_{(l,u)0}$ покидает сеть ($\sum\limits_{k=1}^{N}\sum\limits_{v=1}^{M} p_{(l,u)(k,v)} + p_{(l,u)0} = 1,\ l,k = \overline{1,N};\ u,v = \overline{1,M}$).

В l-м узле находится единственный прибор, который может работать в $r_l + 1$ режимах: 0, 1, ..., r_l, $l = \overline{1,N}$. По истечении времени пребывания в режиме прибор переходит в другой режим мгновенно.

Дисциплина обслуживания заявок прибором – LCFS PR. Заявка, поступающая в узел, вытесняет заявку с прибора и начинает обслуживаться, а вытесненная заявка становится в начало очереди на обслуживание, сдвигая стоящие в ней заявки. При повторном поступлении на прибор требование продолжает дообслуживаться оставшееся время в режиме, в котором работал прибор на момент указанного поступления. Таким образом, поступающая в узел заявка имеет абсолютный приоритет перед всеми остальными заявками, находящимися в узле. Нумерация заявок в очереди на каждый узел осуществляется от конца очереди к прибору.

Состояние сети в момент времени t будем характеризовать вектором $x(t) = (x_1(t), x_2(t), ..., x_N(t))$, где состояние l-го узла в момент времени t описывается вектором $x_l(t) = (\bar{x}_l(t), j_l(t)) = (x_{l1}(t), x_{l2}(t), ..., x_{l,n(l)}(t), j_l(t))$, $x_{l1}(t)$ – тип заявки, стоящей последней в очереди на обслуживание в l-м узле в момент времени t, $x_{l2}(t)$ – тип заявки, стоящей предпоследней в очереди на обслуживание в l-м узле в

момент времени t, и т.д., $x_{l,n(l)-1}(t)$ – тип заявки, стоящей первой в очереди на облуживание в l-м узле в момент времени t, $x_{l,n(l)}(t)$ – тип заявки, находящейся на облуживании в l-м узле в момент времени t, $j_l(t)$ – номер режима, в котором работает прибор в l-м узле в момент времени t, $n(l)$ – общее число заявок в l-м узле. Процесс $x(t)$ имеет пространство состояний $X = X_1 \times X_2 \times ... \times X_N$, где $X_l = \left\{(0, j_l), (x_{l1}, j_l), (x_{l1}, x_{l2}, j_l), ... : x_{lk} = \overline{1, M}, k = 1, 2, ...; j_l = \overline{0, r_l}\right\}$.

Время пребывания в основном (нулевом) режиме работы имеет произвольную функцию распределения $\Phi_l(0, \tilde{u})$, после чего прибор переходит в режим 1. Для состояний x_l, у которых $1 \leq j_l \leq r_l - 1$, время пребывания в режиме j_l также имеет произвольную функцию распределения $\Phi_l(j_l, \tilde{u})$, при этом с вероятностью $\frac{v_l(j_l)}{v_l(j_l)+\varphi_l(j_l)}$ прибор l-го узла переходит в режим $j_l + 1$, а с вероятностью $\frac{\varphi_l(j_l)}{v_l(j_l)+\varphi_l(j_l)}$ – в режим $j_l - 1$. Время пребывания в последнем r_l-м режиме имеет произвольную функцию распределения $\Phi_l(r_l, \tilde{u})$, после чего прибор переходит в режим $r_l - 1$. Во время переключения прибора с одного режима на другой число заявок в узле не меняется.

При этом

$$v_l^{-1}(0) = \int\limits_0^\infty (1 - \Phi_l(0, \tilde{u}))\, d\tilde{u}, \tag{1}$$

$$(v_l(j_l) + \varphi_l(j_l))^{-1} = \int\limits_0^\infty (1 - \Phi_l(j_l, \tilde{u}))\, d\tilde{u},\ 1 \leq j_l \leq r_l - 1, \tag{2}$$

$$\varphi_l^{-1}(r_l) = \int\limits_0^\infty (1 - \Phi_l(r_l, \tilde{u}))\, d\tilde{u}. \tag{3}$$

Длительности обслуживания заявок в узлах не зависят от процесса поступления в случае открытой сети, друг от друга и от длительностей пребывания в режимах и имеют произвольную функцию распределения $B_l(\bar{x}_l, \tilde{u})$ для l-го узла, зависящую от очереди заявок $\bar{x}_l$ в узле, причём

$$\mu_l^{-1}(x_{l1}, x_{l2}, ..., x_{l,n(l)}) = \int\limits_0^\infty \left(1 - B_l(x_{l1}, x_{l2}, ..., x_{l,n(l)}, \tilde{u})\right) d\tilde{u}. \tag{4}$$

Будем предполагать, что матрица $(p_{(l,u)(k,v)}),\ u, v = \overline{1, M},\ l, k = \overline{0, N},\ p_{(0,u)(0,v)} = 0$, неприводима.

Тогда уравнение трафика

$$\varepsilon_{lu} = p_{0(l,u)} + \sum_{k=1}^{N} \sum_{v=1}^{M} \varepsilon_{kv} p_{(k,v)(l,u)},\ l = \overline{1, N},\ u = \overline{1, M}, \tag{5}$$

имеет единственное положительное решение $(\varepsilon_{lu};\ l = \overline{1, N},\ u = \overline{1, M})$.

Пусть $\psi_{lk}(t)$ – остаточное время обслуживания с момента t до завершения обслуживания заявки, стоящей в момент времени t на k-й позиции в l-м узле, $\psi_l(t) = (\psi_{l1}(t), \psi_{l2}(t), ..., \psi_{l,n(l)}(t))$, $l = \overline{1,N}$. Пусть $\xi_{lj_l}(t)$ – остаточное время пребывания прибора l-го узла в режиме j_l с момента t до перехода в соседний режим, $\xi(t) = (\xi_{1,j_1(t)}(t), \xi_{2,j_2(t)}(t), ..., \xi_{N,j_N(t)}(t))$. Таким образом, $\frac{\partial \psi_{lk}(t)}{\partial t} = -\mu_l(x_{l1}, x_{l2}, \ldots, x_{lk})$, $\frac{\partial \xi_{lj_l}(t)}{\partial t} = -\left(v_l(j_l)I_{(j_l \neq r_l)} + \varphi_l(j_l)I_{(j_l \neq 0)}\right)$, когда состояние l-го узла есть $(x_{l1}, x_{l2}, \ldots, x_{l,n(l)}, j_l)$.

В общем случае процесс $x(t)$ не является марковским, поэтому рассмотрим кусочно-линейный марковский процесс $\zeta(t) = (x(t), \psi(t), \xi(t))$, полученный путем добавления к $x(t)$ непрерывных компонент $\psi(t) = (\psi_1(t),\ \psi_2(t),\ ...,\ \psi_N(t))$ и $\xi(t)$.

Под $P = \{P(x)\}$ будем понимать стационарное распределение вероятностей состояний процесса $x(t)$.

Введем в рассмотрение стационарные функции распределения вероятностей состояний кусочно-линейного процесса $\zeta(t)$

$$\begin{aligned} F(x,y,z) &= F(x,\ y_{11},\ y_{12}, ..., y_{1,n(1)}; ...;\ y_{N1},\ y_{N2}, ..., y_{N,n(N)};\ z_{1,j_1}, ..., z_{N,j_N}) = \\ &= \lim_{t\to\infty} P\left\{x(t) = x; \psi_{l1}(t) < y_{l1}, ..., \psi_{l,n(l)}(t) < y_{l,n(l)}, l = \overline{1,N};\right. \\ &\qquad \left.\xi_{1,j_1(t)}(t) < z_{1,j_1}, ..., \xi_{N,j_N(t)}(t) < z_{N,j_N}\right\} \end{aligned} \tag{6}$$

Обозначим $\vartheta_l(j_l) = v_l(j_l)I_{(j_l \neq r_l)} + \varphi_l(j_l)I_{(j_l \neq 0)}$, $l = \overline{1,N}$, $j_l = \overline{0,r_l}$.

3. ОСНОВНОЙ РЕЗУЛЬТАТ

Пусть длительности обслуживания заявок в узлах и длительности пребывания приборов в режимах имеют показательное распределение, т.е. для l-го узла, $l = \overline{1,N}$, $B_l(\bar{x}_l, \tilde{u}) = 1 - \mathbf{e}^{-\mu_l(\bar{x}_l)\tilde{u}}$ и $\Phi_l(j_l, \tilde{u}) = 1 - \mathbf{e}^{-(v_l(j_l)+\varphi_l(j_l))\tilde{u}}$, $(\tilde{u} > 0)$. Тогда $x(t)$– однородный марковский процесс с непрерывным временем и не более чем счетным фазовым пространством состояний $X = X_1 \times X_2 \times ... \times X_N$, где $X_l = \left\{(0, j_l), (x_{l1}, j_l), (x_{l1}, x_{l2}, j_l), (x_{l1}, x_{l2}, x_{l3}, j_l),\ ... : x_{lk} = \overline{1,M},\ k = 1, 2,\ ...;\ j_l = \overline{0, r_l}\right\}$.

Из работы [2] следует, что если положить $v_l(\bar{x}_l, j_l) = v_l(j_l)$, $\varphi_l(\bar{x}_l, j_l) = \varphi_l(j_l)$, то при выполнении условия

$$\sum_{x\in X} q(x) \prod_{l=1}^{N} \left[\lambda^{n(l)} \prod_{w=1}^{n(l)} \frac{\varepsilon_{l,x_{lw}}}{\mu_l(x_{l1}, x_{l2}, ..., x_{lw})} \prod_{k=1}^{j_l} \frac{v_l(k-1)}{\varphi_l(k)}\right] < \infty, \tag{7}$$

процесс $x(t)$ эргодичен, а его финальное стационарное распределение имеет мультипликативную форму $P(x) = p_1(x_1)p_2(x_2) \times ... \times p_N(x_N)$.

При этом

$$p_l(\bar{x}_l, j_l) = \left(\lambda^{n(l)} \prod_{w=1}^{n(l)} \frac{\varepsilon_{l,x_{lw}}}{\mu_l(x_{l1}, x_{l2}, ..., x_{lw}, j_l)} \prod_{k=1}^{j_l} \frac{v_l(k-1)}{\varphi_l(k)}\right) p_l(0,0),$$

где $q(x) = \lambda + \sum\limits_{l=1}^{N} \left(\mu_l^{-1}(\bar{x}_l) + v_l(j_l) + \varphi_l(j_l)\right)$, а $(\varepsilon_{lu};\ l = \overline{1,N},\ u = \overline{1,M})$ – положительное решение уравнения трафика (5).

Пусть теперь длительность обслуживания заявки прибором l-го узла имеет произвольную функцию распределения $B_l(\bar{x}_l, \tilde{u})$, а время пребывания прибора l-го узла в режиме j_l – произвольную функцию распределения $\Phi_l(j_l, \tilde{u})$, причем математические ожидания фиксированы с помощью равенств (1)-(4).

Основной результат формулируется следующим образом.

Теорема 1. *Процесс $\zeta(t)$ эргодичен, если выполняются соотношения (7). При этом стационарные функции распределения вероятностей состояний $F(x, y, z)$ определяются по формулам*

$$F(x,y,z) = p_1(x_1)p_2(x_2) \times ... \times p_N(x_N) \times$$
$$\times \prod_{l=1}^{N} \prod_{w=1}^{n(l)} \mu_l(x_{l1}, ..., x_{lw}) \int_0^{y_{l,w}} (1 - B_l(x_{l1}, ..., x_{lw}, \tilde{u}))d\tilde{u} \prod_{l=1}^{N} \vartheta_l(j_l) \int_0^{z_{l,j_l}} (1 - \Phi_l(j_l, \tilde{u}))d\tilde{u}, \tag{8}$$

где

$$p_l(\bar{x}_l, j_l) = \left(\lambda^{n(l)} \prod_{w=1}^{n(l)} \frac{\varepsilon_{l,x_{lw}}}{\mu_l(x_{l1}, ..., x_{lw})} \prod_{k=1}^{j_l} \frac{v_l(k-1)}{\varphi_l(k)} \right) p_l(0,0), \tag{9}$$

$\varepsilon_{l,x_{lw}}$ находятся из (5), а

$$p_l(0,0) = \left(\sum_{i=0}^{\infty} \sum_{j=0}^{r_l} \lambda^i \prod_{w=1}^{i} \frac{\varepsilon_{l,x_{lw}}}{\mu_l(x_{l1}, ..., x_{lw})} \prod_{k=1}^{j} \frac{v_l(k-1)}{\varphi_l(k)} \right)^{-1}, \ x \in X, \ l = \overline{1, N}. \tag{10}$$

Из теоремы 1 с учетом равенства $P(x) = F(x, +\infty, +\infty)$ получаем следующее утверждение.

Следствие. *Если выполняются соотношения (7), то процесс $x(t)$ эргодичен, а его стационарное распределение $P = \{P(x), \ x \in X\}$ не зависит от функционального вида распределений $B_l(\bar{x}_l, \tilde{u})$, $\Phi_l(k, \tilde{u})$ и имеет вид $P(x) = p_1(x_1)p_2(x_2) \times ... \times p_N(x_N)$, где $p_l(x_l)$ определяются по формулам (9)-(10).*

ЛИТЕРАТУРА

1. *Малинковский Ю. В., Нуеман А. Ю.* Мультипликативность стационарного распределения в открытых сетях с многорежимными стратегиями обслуживания // Весцi НАН Беларусi. 2001. № 3. С. 129–134.
2. *Летунович Ю. Е.* Стационарное распределение состояний открытой неоднородной сети с многорежимными стратегиями и немедленным обслуживанием // Современные информационные компьютерные технологии: сб. науч. ст. в 2 ч (ч. 2). Гродно, 2008. С. 97–99.
3. *Малинковский Ю. В., Старовойтов А. Н., Ерёмина А. Р.* Инвариантность стационарного распределения вероятностей состояний сетей с многорежимными стратегиями обслуживания, разнотипными заявками и дисциплиной обслуживания LCFS PR // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2009. № 3(8). С. 33–39.

ПОСТРОЕНИЕ И ИССЛЕДОВАНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ НЕОДНОРОДНОГО ДОРОЖНОГО ТРАФИКА

М. Федоткин*, **Е. Кудрявцев**

Нижегородский госуниверситет им. Лобачевского
национальный исследовательский университет
г. Нижний Новгород, Россия
*fma5@rambler.ru

В работе изучена случайная последовательность моментов пересечения автомобилями некоторой поперечной линии магистрали при плохих погодных и дорожных условиях. Интервалы между такими моментами являются зависимыми и имеют различные распределения. Рассмотрен нетрадиционный способ описания потока машин такой сложной вероятностной структуры. Этот способ основан на изучении распределении величины транспортной пачки и на распределении потока, так называемых, медленных или головных в автоколонне машин. При этом предполагается интенсивное движение другого типа машин – быстрых машин.

Ключевые слова: транспортная пачка, поток Пуассона, бесконечная система дифференциальных уравнений Колмогорова, предельное распределение.

1. ВВЕДЕНИЕ

Математической теории транспортных потоков посвящено большое число монографий и статей [1]. При этом предполагается, что транспортный поток сотоит из однородных или однотипных машин. Для реального транспортного потока машин это ограничение, как правило, не выполняется. Более того, каждый автомобиль осуществляет движение в экстремальных погодных и дорожных условиях, скорость которого является случайной величиной. В силу этого транспортный поток машин на магистралях существенно отличается от потока случайных событий, который рассматривается в классической теории массового обслуживания. Например, всегда требуется определить эргодическое распределение случайного числа машин в транспортной пачке. Для транспортного потока, как правило, не удается изучить вероятностные свойства последовательности из случайных моментов пересечения машинами, так называемой, виртуальной стоп-линии. В этой работе рассматриваются вероятностные модели транспортных потоков с учетом пространственных и временных свойств процессов движения автомобилей на автомагистрали. На реальных примерах показана эффективность предложенного подхода при построении и изучении моделей транспортных потоков.

2. РАСПРЕДЕЛЕНИЕ ТРАНСПОРТНОЙ ПАЧКИ

При удовлетворительном состоянии дорожного полотна и хороших метеорологических условиях движение разного типа автомобилей по магистрали может оказаться беспрепятственным и пуассоновским. При плохих погодных условиях (туман, снег, гололед и т. д.) обгон быстрыми машинами медленных является уже рискованным, зависимым и занимает значительное время. В этом случае на интенсивных магистралях будут возникать автоколонны (группы) машин, или транспортные пачки, т. е. транспортные потоки уже не будут пуассоновскими. С такой ситуацией впервые столкнулся в 1963 году Бартлетт [2] при наблюдении за движением машин вблизи Лондона. Для такого типа транспортного потока Бартлетту и другим исследователям не удалось найти подходящего закона распределения для зависимых промежутков времени между двумя последовательными пересечениями автомобилями виртуальной стоп-линии. Один из авторов данной работы приблизительно в это же время решал задачу [3] об оптимальном управлении транспортными потоками в городе Горьком (Нижний Новгород). По наблюдениям за движением машин на магистралях вблизи Горького и других крупных городов было подмечено, что транспортная пачка состоит из головной машины с медленным движением и очереди из быстрых машин, которые догнали медленную и ожидают возможности обгона. Здесь возникает проблема построения модели пространственного расположения машин на магистрали. С этой целью был предложен простой механизм образования транспортных пачек при интенсивном движении машин в плохих погодных условиях. Для большого числа магистралей оказалось, что быстрые машины поступают в транспортную пачку, или, другими словами, догоняют медленную машину, по закону Пуассона с интенсивностью $\lambda > 0$. Это означает, что быстрые машины осуществляют относительно свободное движение на протяженных участках дороги, где нет медленных машин. Подробно такой транспортный поток был изучен в задаче о свободном движении автомобилей по магистрали [1]. Если $(\Omega, \mathcal{F}, \mathbf{P}(\cdot))$ — основное вероятностное пространство, то далее через ω обозначим произвольный элемент достоверного события Ω. В некоторых случаях символ ω будем опускать. В соответствии с этой задачей обозначим через $\xi(\omega; t)$ случайное число быстрых машин, которые поступают по закону Пуассона в транспортную пачку за промежуток времени $[0, t)$. Аналогично обозначим через $\xi(\omega; t, \Delta t)$ случайное число быстрых машин, которые поступают по закону Пуассона в транспортную пачку за промежуток времени $[t, t + \Delta t)$.

Каждую машину с медленным движением можно интерпретировать как обслуживающий прибор для машин с быстрым движением. При этом под временем обслуживания, естественно, понимается случайное время обгона. На практике среднее время обслуживания (обгона) существенно зависит от числа машин в транспортной пачке. Пусть случайная величина $\chi(\omega; t)$ измеряет число всех типов машин в транспортной пачке в момент времени $t \geq 0$. Обозначим теперь через $\eta(\omega; t, \Delta t)$ случайное число быстрых машин, которые могу обогнать медленную за промежуток времени $[t, t + \Delta t)$. Вполне естественно предположить, что при малых значениях $\Delta t > 0$ условные вероятности событий, которые порождаются дискрет-

ной случайной величиной $\eta(\omega; t, \Delta t)$, определяются следующими соотношениями:

$$\begin{gathered}
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=0\}|\{\omega\colon \chi(\omega;t)=2, \xi(\omega;t,\Delta t)=0\}) = 1-\mu_1\Delta t + o(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=1\}|\{\omega\colon \chi(\omega;t)=2, \xi(\omega;t,\Delta t)=0\}) = \mu_1\Delta t - o(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=0\}|\{\omega\colon \chi(\omega;t)=3, \xi(\omega;t,\Delta t)=0\}) = 1-\mu_2\Delta t + o(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=1\}|\{\omega\colon \chi(\omega;t)=3, \xi(\omega;t,\Delta t)=0\}) = \mu_2\Delta t - o(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=0\}|\{\omega\colon \chi(\omega;t)=m, \xi(\omega;t,\Delta t)=0\}) = 1-\mu_3\Delta t + o(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=1\}|\{\omega\colon \chi(\omega;t)=m, \xi(\omega;t,\Delta t)=0\}) = \mu_3\Delta t - o(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=0\}|\{\omega\colon \chi(\omega;t)=m-3, \xi(\omega;t,\Delta t)=1\}) = 1-O(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)=1\}|\{\omega\colon \chi(\omega;t)=m-3, \xi(\omega;t,\Delta t)=1\}) = O(\Delta t),\\
\mathbf{P}(\{\omega\colon \eta(\omega;t,\Delta t)\geq 2\}|\{\omega\colon \chi(\omega;t)=m-2, \xi(\omega;t,\Delta t)=1\}) = o(\Delta t),\\
m=4,5,\ldots
\end{gathered} \tag{1}$$

Здесь и далее $o(\Delta t)$ есть бесконечно малая неотрицательная величина по сравнению с Δt при $t \to 0$. В равенствах (1) параметры μ_1^{-1} и μ_2^{-1} задают среднее время обгона каждой быстрой машиной медленную в случае, когда транспортная пачка состоит из двух и трех машин соответственно. Аналогично параметр μ_3^{-1} в равенствах (1) определяет среднее время обгона каждой быстрой машиной медленную, если транспортная пачка состоит из четырех и более машин. Параметры μ_1, μ_2, μ_3 будем называть интенсивностями обгона. Таким способом моделируется зависимость среднего времени обгона от числа машин в транспортной пачке. Вероятности в (1) не зависят от времени t. Поэтому в дальнейшем ради упрощения записи будем опускать символ t и обозначать величину $\eta(\omega; t, \Delta t)$ через $\eta(\omega; \Delta t)$.

Восьмое и последнее равенства соотношения (1) можно проинтерпретировать следующим образом. При заданном размере транспортной пачки условная вероятность того, что за (где угодно расположенный) промежуток Δt по меньшей мере две машины обгонят медленную, есть величина бесконечно малая по сравнению с Δt. Обозначим теперь через $Q(t, m)$ вероятность $\mathbf{P}(\{\omega\colon \chi(\omega; t) = m\})$ при фиксированных $t > 0$ и $m = 1, 2, \ldots$ Итак, набор вероятностей $Q(t, m)$, $m = 1, 2, \ldots$ определяет распределение числа машин в транспортной пачке в момент $t \geq 0$.

Рассуждениями, аналогичными тем, которые обычно используются в теории массового обслуживания, мы можем получить бесконечную систему линейных дифференциальных уравнений первого порядка для вероятностей вида $Q(t, m)$, $t \geq 0$, $m = 1, 2, \ldots$ При каждом фиксированном $m = 1, 2, \ldots$ для случайного события $\{\omega\colon \chi(\omega; t + \Delta t) = m\}$ нетрудно показать следующее равенство в событиях

$$\{\omega\colon \chi(\omega;t+\Delta t)=m\} = \bigcup_{k=1}^{\infty}\bigcup_{n=0}^{\infty}\{\omega\colon \chi(\omega;\Delta t)=k, \xi(\omega;\Delta t)=n, \eta(\omega;\Delta t)=k+n-m\}.$$

Используя это соотношение и формулы (1), легко получить бесконечную систему линейных дифференциальных уравнений Колмогорова:

$$
\begin{aligned}
dQ(t,1)/dt &= -\lambda Q(t,1) + \mu_1 Q(t,2),\\
dQ(t,2)/dt &= \lambda Q(t,1) - (\lambda+\mu_1)Q(t,2) + \mu_2 Q(t,3),\\
dQ(t,3)/dt &= \lambda Q(t,2) - (\lambda+\mu_2)Q(t,3) + \mu_3 Q(t,4),\\
dQ(t,m)/dt &= \lambda Q(t,m-1) - (\lambda+\mu_3)Q(t,m) + \mu_3 Q(t,m+1), m \geq 4.
\end{aligned} \tag{2}
$$

Будем дополнительно предполагать, что в момент $t = 0$ число машин в транспортной пачке равно i. Тогда динамика распределения числа машин в транспортной пачке определяется решением системы дифференциальных уравнений (2) с начальными условиями $Q(0,i) = 1$, $Q(0,m) = 0$ при $m \geq 1$ и $m \neq i$.

Явное решение системы дифференциальных уравнений (2) может быть найдено с помощью вывода и последующего решения дифференциального уравнения в частных производных для производящей функции $\Pi_{\chi(t)}(t,z) = \sum_{m=1}^{\infty} z^m Q(t,m)$, как это делается в большинстве задач теории массового обслуживания. Однако в этом случае процесс решения является очень громоздким и далеко нетривиальным. К счастью, нам потребуется здесь только свойства распределения числа машин в транспортной пачке при $t \to \infty$, т. е. некоторые свойства решений системы (2) при $t \to \infty$. Общие свойства решений такого рода систем дифференциальных уравнений детально изучены в работах Колмогорова, Феллера, Ледермана, Карлина, Кларка, Мак-Грегора, Рейтера, Федоткина. В частности, если $\lambda < \mu_3$, то существует единственное решение системы уравнений (2), удовлетворяющее при $m = 1, 2, \ldots$ условиям: $\sum_{m=1}^{\infty} Q(t,m) = 1, \lim_{t\to\infty} Q(t,m) = Q(m) > 0, \sum_{m=1}^{\infty} Q(m) = 1$. При этом указанные пределы не зависят от начальных условий и могут быть получены путем решения бесконечной системы линейных алгебраических уравнений

$$
\begin{gathered}
0 = -\lambda Q(1) + \mu_1 Q(2), \quad 0 = \lambda Q(1) - (\lambda+\mu_1)Q(2) + \mu_2 Q(3),\\
0 = \lambda Q(2) - (\lambda+\mu_2)Q(3) + \mu_3 Q(4),\\
0 = \lambda Q(m-1) - (\lambda+\mu_3)Q(m) + \mu_3 Q(m+1), \quad m = 4,\ 5,\ \ldots
\end{gathered} \tag{3}
$$

Распределение $\{Q(m); m = 1, 2, \ldots\}$ называется предельным или эргодическим для числа $\chi(\omega)$ всех типов машин в транспортной пачке. Это распределение характеризует так называемый установившийся или стационарный режим движения автоколонн по магистрали. На содержательном уровне условие $\lambda < \mu_3$ означает, что интенсивность, с которой быстрые машины догоняют медленную, должна быть меньше интенсивности обгона. При $\lambda < \mu_3$ система (3) получается с помощью предельного перехода при $t \to \infty$ одновременно во всех уравнениях системы (2) с учетом равенств $\lim_{t\to\infty}(dQ(t,m)/dt) = 0, \lim_{t\to\infty} Q(t,m) = Q(m), m \geq 1$. Отметим, что при каждом фиксированном $m \geq 1$ предел $\lim_{t\to\infty}(dQ(t,m)/dt)$ существует. Более того, каждый такой предел равен нулю. В противном случае модуль величины $Q(t,m)$ при $t \to \infty$ возрастал бы неограниченно, что невозможно в силу смысла величины $Q(t,m)$ как вероятности. Перейдем теперь к решению системы (3).

Пусть при $m = 3, 4, \ldots$ величина $u_m = -\lambda Q(m) + \mu_3 Q(m+1)$. Из первых трех уравнений системы (3) получим, что $Q(2) = \lambda\mu_1^{-1}Q(1)$, $Q(3) = \lambda^2\mu_1^{-1}\mu_2^{-1}Q(1)$ и

$Q(4) = \lambda^3\mu_1^{-1}\mu_2^{-1}\mu_3^{-1}Q(1)$. Поэтому $u_3 = -\lambda Q(3) + \mu_3 Q(4) = 0$. Из системы (3) найдем $u_m - u_{m-1} = 0, m \geq 4$. Отсюда $u_m = 0$ для $m \geq 3$. Значит, для $m \geq 3$ вероятность $Q(m) = \lambda^2\mu_1^{-1}\mu_2^{-1}(\lambda\mu_3^{-1})^{m-3}Q(1)$. Замена $\alpha = \lambda\mu_1^{-1}, \beta = \lambda\mu_2^{-1}, \gamma = \lambda\mu_3^{-1}$, условие $\sum_{m=1}^{\infty} Q(m) = 1$ и неравенство $\lambda < \mu_3$ окончательно дает:

$$\begin{gathered} Q(1) = (1+\alpha+\alpha\beta/(1-\gamma))^{-1}, Q(2) = \alpha(1+\alpha+\alpha\beta/(1-\gamma))^{-1}, \\ Q(m) = \alpha\beta\gamma^{m-3}(1+\alpha+\alpha\beta/(1-\gamma))^{-1}, m \geq 3. \end{gathered} \tag{4}$$

Производящая функция $\Pi_\chi(t,z) = \sum_{m=1}^{\infty} z^m Q(m) = p(z+\alpha z^2+\alpha\beta z^3(1-\gamma z)^{-1})$, где $p = (1+\alpha+\alpha\beta/(1-\gamma))^{-1}$. Отсюда $M\chi(\omega) = p(1+2\alpha+\alpha\beta[2(1-\gamma)^{-1}+(1-\gamma)^{-2}])$, $D\chi(\omega) = (\alpha+\alpha\beta[(1-\gamma)^{-1}+(1-\gamma)^{-2}+2(1-\gamma)^{-3}]+\alpha^2\beta[-(1-\gamma)^{-2}+2(1-\gamma)^{-3}])p^2+$ $+p^2\alpha^2\beta^2[-(1-\gamma)^{-3}+(1-\gamma)^{-4}]$. Вид формулы (4) и равенство $\gamma = \lambda\mu_3^{-1}$ позволяет дать следующий простой смысл параметра γ для задачи о групповом движении машин по магистрали. Параметр γ задает степень насыщения транспортной магистрали быстрыми машинами и его можно назвать коэффициентом заполнения или загрузкой транспортной магистрали. В самом деле, если $\mu_3 > \lambda$ и $\mu_3 \to \lambda$, то легко выводим, что параметры $p \to 0, \gamma \to 1$. Поэтому при $\mu_3 > \lambda$ и $\mu_3 \to \lambda$ математическое ожидание числа всех типов машин в транспортной пачке для стационарного режима растет неограниченно. Это обстоятельство постоянно наблюдают автомобилисты при существенном ухудшении погодных условий, когда на дорогах образуются транспортные заторы значительной протяженности. Теперь можно сделать одно важное замечание. Легко видеть, что при $\alpha = \beta = \gamma$ данное распределение совпадает с геометрическим распределением. Из равенств $\alpha = \lambda\mu_1^{-1}, \beta = \lambda\mu_2^{-1}, \gamma = \lambda\mu_3^{-1}$ находим, что это возможно только при $\mu_1 = \mu_2 = \mu_3$. Другими словами, если среднее время обгона каждой быстрой машиной медленную не зависит от числа машин в транспортной пачке, то мы имеем дело с геометрическим законом распределения.

3. СВОЙСТВА НЕОРДИНАРНОГО ПОТОКА

В реальном транспортном потоке плотность медленных машин (число медленных машин на единичном по длине участке автомагистрали) значительно меньше плотности быстрых. В тоже время можно допустить, что движение медленных машин в стационарном режиме происходит независимым образом. В связи с этим можно предположить, что транспортный поток из медленных машин в стационарном режиме будет пуассоновским с интенсивностью μ. Пусть максимальная длина участка дороги, на котором располагается среднее число $M\chi(\omega)$ автомобилей в транспортной пачке, много меньше среднего расстояния между соседними автомобилями с медленным движением. Тогда можно считать, что все машины каждой движущейся автоколонны в стационарном режиме пересекают некоторую поперечную линию автомагистрали одновременно. Для данного потока обозначим теперь через $\varkappa(\omega;t)$ случайное число всех типов машин, которые пересекают фиксированную поперечную линию автомагистрали за промежуток времени $[0,t)$. Для $\varkappa(\omega;t)$ введем производящую функцию $\Pi_{\varkappa(t)}(t,z) = \sum_{m=0}^{\infty} z^m \mathbf{P}(\{\omega\colon \varkappa(\omega;t) = m\})$.

Теорема 1. *Для функции* $\Pi_{\varkappa(t)}(t,z)$ *справедливо равенство* $\Pi_{\varkappa(t)}(t,z) = e^{-\mu t} \times \times \sum_{k=0}^{\infty} z^k \left\{ \sum_{n=0}^{[\frac{k}{2}]} \alpha^n \left[\frac{(\mu tp)^{k-n}}{n!(k-2n)!} + \sum_{m=1}^{min\{k-2n,n\}} \beta^m \sum_{l=0}^{k-2n-m} \gamma^l \frac{(\mu tp)^{k-n-m-l} C_{m+l-1}^l}{(n-m)!m!(k-2n-m-l)!} \right] \right\}$, *где* $\left[\frac{k}{2}\right]$ *обозначает целую часть числа* $\frac{k}{2}$.

Из теоремы следует, что поток определяется параметрами μ, α, β и γ.

Теорема 2. *Для распределения* $\mathbf{P}(\{\omega\colon \varkappa(\omega;t) = k\})$, $k \geq 0$ *случайного потока* $\{\varkappa(\omega;t)\colon t \geq 0\}$ *имеет место:* $\mathbf{P}(\{\omega\colon \varkappa(\omega;t) = 0\}) = e^{-\mu t}, \mathbf{P}(\{\omega : \varkappa(\omega;t) = k\}) =$

$$= e^{-\mu t} \sum_{n=0}^{[\frac{k}{2}]} \alpha^n \left[\frac{(\mu tp)^{k-n}}{n!(k-2n)!} + \sum_{m=1}^{min\{k-2n,n\}} \beta^m \sum_{l=0}^{k-2n-m} \gamma^l \frac{(\mu tp)^{k-n-m-l} C_{m+l-1}^l}{(n-m)!m!(k-2n-m-l)!} \right].$$

Лемма 1. *Пусть* $M\varkappa(\omega;t)$ *и* $D\varkappa(\omega;t)$ *суть математическое ожидание и дисперсия* $\varkappa(\omega;t)$. *Тогда имеет место:* $M\varkappa(\omega;t) = \mu tp \left(1 + 2\alpha + \alpha\beta \left[\frac{2}{1-\gamma} + \frac{1}{(1-\gamma)^2}\right]\right)$, $D\varkappa(\omega;t) = \mu tp \left(1 + 4\alpha + \alpha\beta \left[\frac{4}{1-\gamma} + \frac{3}{(1-\gamma)^2} + \frac{2}{(1-\gamma)^3}\right]\right)$.

Лемма 2. *Для того чтобы сумма* n *независимых потоков данного вида с параметрами* μ_i, α_i, β_i *и* γ_i, $i = \overline{1,n}$, *являлась таким же потоком, необходимо и достаточно, чтобы* $\gamma_1 = \gamma_2 = ... = \gamma_n$. *При этом суммарный поток имеет параметры:* $\mu = \sum_{i=1}^n \mu_i, \alpha = \sum_{i=1}^n \mu_i\alpha_i p_i / \sum_{i=1}^n \mu_i p_i, \beta = \sum_{i=1}^n \mu_i\alpha_i\beta_i p_i / \sum_{i=1}^n \mu_i\alpha_i p_i, \gamma = \gamma_1$.

4. ЗАКЛЮЧЕНИЕ

Была решена проблема построения и изучения математической модели пространственной и временной характеристик неоднородного транспортного потока на магистрали при большой плотности быстрых машин и значительном расстоянии между последовательными медленными машинами. При этом для построения вероятностной модели пространственной характеристики транспортного потока используется нелокальный способ [3] описания потоков неоднородных заявок. Показано, что локальное описание временной характеристики стационарного движения транспортного потока неоднородных машин можно выполнить неординарным пуассоновским потоком, когда в каждый вызывающий момент с вероятностью единица поступает лишь конечное число автомобилей. Подробно изучены вероятностные свойства и свойства числовых характеристик такого потока.

ЛИТЕРАТУРА

1. *Haight F. A.* Mathematical theories of traffic flow. New York London: Academic press, 1963.
2. *Bartlett M. S.* The spectral analysis of point processes // J. R. Statist. Soc. B. 1963. V. 25. № 2. P. 264–296.
3. *Федоткин М. А.* Нелокальный способ задания управляемых случайных процессов // Математические вопросы кибернетики, 1998. № 7. С. 332–344.

INVESTIGATION OF TRAFFIC FLOWS CHARACTERISTICS IN CASE OF THE SMALL DENSITY

M. Fedotkin*, **M. Rachinskaya**

N. I. Lobachevsky State University of Nizhni Novgorod
National Research University
Nizhni Novgorod, Russia
*fma5@rambler.ru

A random number of vehicles crossed a transverse line of a motorway during an arbitrary interval of time and a random number of vehicles situated on an arbitrary part of the motorway at a fixed instant of time form a complicated stochastic dependence. This is indicated by, for example, a functional dependence between the intensity of the traffic and its density[1]. First mathematical model of traffic flow of non-homogeneous vehicles in case of bad weather and bad road conditions is constructed taking into account both spatial and temporal processes.

Keywords: traffic flow, Poisson flow, finite system of Kolmogorov differential equations, limiting probability distribution.

1. INTRODUCTION

There are usually considered such traffic flows in the classical queueing theory, random distances between the neighbouring vehicles in which are independent and all of them have exponential probability distributions. However in practice we often encounter with a situation, when the intervals between moments of crossing a virtual transverse line of the motorway by the consecutive vehicles are dependent and have different probability distributions. In this case constructing and investigation of the models of spacial and temporal traffic flows characteristics offer some significant difficulties. This paper proposes simple mechanism of so-called traffic batch formation. It makes constructing and researching of vehicles spacial location on the motorway possible. Assuming it as a basis we managed to find and ground simple local description of temporal characteristic of stationary traffic flow of non-homogeneous vehicles. By this token it is proved that we can use non-ordinary Poisson process to describe adequately the traffic on the motorway taking into account both its spatial and temporal characteristics. Meanwhile a number of arrivals received in a random manner at every calling instant of crossing a transverse line of the motorway by the vehicles is restricted.

2. MECHANISM OF TRAFFIC BATCH FORMATION AND ITS PROBABILITY PROPERTIES

The effect of some factors, such as bad weather conditions, heterogeneity of vehicles or unsatisfactory conditions of the road surface, make difficulties for the unimpeded motorway traffic. Under these conditions vehicles aren't able to overtake each other freely. So we can observe foundation of traffic groups (batches, motorcades). It doesn't allow us to consider traffic flow as the Poisson flow. Here the necessity of researching of the traffic flow spacial characteristic appears. First the spacial model of traffic flow of homogeneous vehicles was investigated with the usage of the Poisson process in [1]. Basing on the observation over the motorway traffic close to Gorky town [2] the following mechanism of traffic batch formation was proposed.

Every motorcade consists of one slow vehicle at the head of it and some fast vehicles that wait for the possibility of overtaking. That's why traffic batch was proposed to be considered as a service system with varying structure. So every slow vehicle can be interpreted as an servicing device for the fast vehicles. Here the service means overtaking slow vehicle by the fast one. We can take notice, that without slow vehicles fast vehicles move unimpeded enough to suppose their inflow in the motorcade having Poisson distribution. We will designate a random number of vehicles flowed into the the traffic batch for interval of time $[0, t)$ following the Poisson law with an parameter λ_0 as $\eta_0(\omega; t)$. We will consider basic probability space $(\Omega, \mathfrak{F}, \mathbf{P}(\cdot))$ below, where ω means elementary outcome or an element of the certain event Ω. Further we will designate as $\xi(\omega; t, \Delta t)$ a random number of fast vehicles overtook slow one for the interval of time $[t, t + \Delta t)$ and introduce a random variable $\varkappa(\omega; t)$ which counts a number of all vehicles in a traffic batch at the instant of time $t \geq 0$. Let $\varkappa(\omega; t)$ accept the values from the set $\{1, 2, \cdots, N\}$. It is possible if the intensity of overtaking slow vehicles by fast exceeds the intensity of the fast vehicles inflow into the motorcade considerably. In this case relatively small motorcade generates indeed. As far as the average overtaking time depends on a number of vehicles in the batch it is necessary to discriminate the following situations. Let μ_1^{-1} and μ_2^{-1} be an average overtaking time in case when the batch consists of two and three vehicles, respectively. Also let us assume that average time doesn't change if there more than three vehicles in the motorcade and we will designate it as μ_3^{-1}. Let the parameters μ_1, μ_2 and μ_3 term the overtaking intensities in cases mentioned above. We should impose constraint $\lambda_0 < \mu_3$ on the system parameters. This constraint means intensionally such condition when overtaking intensity exceeds the intensity of the vehicles inflow into the motorcade. Taking into account what the vehicles shouldn't be lost and what the batch can't consist of more than $N \geq 4$ vehicles we conjecture the fulfilment of the next constraint. If the fast vehicle catches up with the full batch of N vehicles, it will join the batch nevertheless. But at the same time the fast vehicle following the slow one will certainly overtake it. Now we can define in case of small $\Delta t > 0$ the conditional probability of events generated by the random

variable $\xi(\omega;t,\Delta t)$ with the help of the following relations:

$$\begin{gathered}
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=0\}|\{\omega\colon \varkappa(\omega;t)=1,\eta_0(\omega;t,\Delta t)=1\})=1-O(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=0\}|\{\omega\colon \varkappa(\omega;t)=2,\eta_0(\omega;t,\Delta t)=0\})=1-\mu_1\Delta t+o(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=1\}|\{\omega\colon \varkappa(\omega;t)=2,\eta_0(\omega;t,\Delta t)=0\})=\mu_1\Delta t-o(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=0\}|\{\omega\colon \varkappa(\omega;t)=3,\eta_0(\omega;t,\Delta t)=0\})=1-\mu_2\Delta t+o(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=1\}|\{\omega\colon \varkappa(\omega;t)=3,\eta_0(\omega;t,\Delta t)=0\})=\mu_2\Delta t-o(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=0\}|\{\omega\colon \varkappa(\omega;t)=k,\eta_0(\omega;t,\Delta t)=0\})=1-\mu_3\Delta t+o(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=1\}|\{\omega\colon \varkappa(\omega;t)=k,\eta_0(\omega;t,\Delta t)=0\})=\mu_3\Delta t-o(\Delta t),\\
\mathbf{P}(\{\omega\colon \xi(\omega;t,\Delta t)=1\}|\{\omega\colon \varkappa(\omega;t)=N,\eta_0(\omega;t,\Delta t)=1\})=1,
\end{gathered}\tag{1}$$

where $4\le k\le N$. In the relation (1) symbols $O(\Delta t)$ and $o(\Delta t)$ are infinitesimal of the same order relative Δt and infinitesimal of higher order relative Δt, respectively.

We will designate as $Q(t,k)$ the probability $\mathbf{P}(\{\omega\colon \varkappa(\omega;t)=k\})$ that determines when $k=1,2,\ldots,N$ and $t\ge 0$. It should be noticed what the following equality takes place for every chance event of the form $\{\omega\colon \varkappa(\omega;t+\Delta t)=k\}$, $k=1,2,\ldots,N$:

$$\{\omega\colon \varkappa(\omega;t+\Delta t)=k\}=\bigcup_{l=1}^{N}\bigcup_{m=0}^{\infty}\{\omega\colon \varkappa(\omega;t)=l,\eta_0(\omega;t,\Delta t)=m,\xi(\omega;t,\Delta t)=l+m-k\}.$$

Basing on it and relations derived above we can write down the next equalities:

$$\begin{gathered}
Q(t+\Delta t,1)=(1-\lambda_0\Delta t)Q(t,1)+\mu_1\Delta tQ(t,2)+o(\Delta t),\\
Q(t+\Delta t,2)=\lambda_0\Delta tQ(t,1)+(1-\lambda_0\Delta t-\mu_1\Delta t)Q(t,2)+\mu_2\Delta tQ(t,3)+o(\Delta t),\\
Q(t+\Delta t,3)=\lambda_0\Delta tQ(t,2)+(1-\lambda_0\Delta t-\mu_2\Delta t)Q(t,3)+\mu_3\Delta tQ(t,4)+o(\Delta t),\\
Q(t+\Delta t,k)=\lambda_0\Delta tQ(t,k-1)+(1-\lambda_0\Delta t-\mu_3\Delta t)Q(t,k)+\mu_3\Delta tQ(t,k+1)+o(\Delta t),\\
k=1,2,\ldots,N-1,\\
Q(t+\Delta t,N)=\lambda_0\Delta tQ(t,N-1)+(1-\lambda_0\Delta t-\mu_3\Delta t)Q(t,N)+\lambda_0\Delta tQ(t,N)+o(\Delta t).
\end{gathered}$$

Proceeding to the limit $\Delta t\to 0$ we receive a system of N linear homogeneous Kolmogorov differential first-order equations with the constant coefficients and Jacobi matrix:

$$\begin{gathered}
dQ(t,1)/dt=-\lambda_0Q(t,1)+\mu_1Q(t,2),\\
dQ(t,2)/dt=\lambda_0Q(t,1)-(\lambda_0+\mu_1)Q(t,2)+\mu_2Q(t,3),\\
dQ(t,3)/dt=\lambda_0Q(t,2)-(\lambda_0+\mu_2)Q(t,3)+\mu_3Q(t,4),\\
dQ(t,k)/dt=\lambda_0Q(t,k-1)-(\lambda_0+\mu_3)Q(t,k)+\mu_3Q(t,k+1),4\le k<N,\\
dQ(t,N)/dt=\lambda_0Q(t,N-1)-\mu_3Q(t,N).
\end{gathered}\tag{2}$$

This system describes probability distribution of the traffic batch length dynamics. We can consider the entry conditions of the form $Q(0,k)=\delta_{ik}$, i.e. $Q(0,i)=1$ and $Q(0,k)=0$ where $k\ne i$.

Solution of system (2) can be derived with the help of different methods of differential equations theory. However in general case it is too lengthy. Luckily in our case investigation of only some solution properties with tending $t \to \infty$ is really important. Let's designate $\lim\limits_{t\to\infty} Q(t,k) = Q(k)$ for $k = 1, 2, \ldots, N$. According to Markov theorem these limits exist. Then the limits of functions $dQ(t,k)/dt$ exist too. Moreover all of them are equal to 0. Indeed, if such $j \in \{1, 2, \ldots, N\}$ that $\lim\limits_{t\to\infty} \frac{dQ(t,j)}{dt} \neq 0$ had existed we would have gotten the infinite increase of absolute value of quantity $Q(t,j)$. Such increase is impossible because quantity $Q(t,j)$ defines the probability so it is limited. Relying on these reasonings we can pass from differential equations system (2) to system of linear equations (3):

$$\begin{gathered} 0 = -\lambda_0 Q(1) + \mu_1 Q(2), \quad 0 = \lambda_0 Q(1) - (\lambda_0 + \mu_1) Q(2) + \mu_2 Q(3), \\ 0 = \lambda_0 Q(2) - (\lambda_0 + \mu_2) Q(3) + \mu_3 Q(4), \\ 0 = \lambda_0 Q(k-1) - (\lambda_0 + \mu_3) Q(k) + \mu_3 Q(k+1), \quad k = 4,\ 5,\ \ldots,\ N-1, \\ 0 = \lambda_0 Q(N-1) - \mu_3 Q(N). \end{gathered} \tag{3}$$

This system determines so-called ergodic probability distribution $\{Q(k); 1 \leq k \leq N\}$. It characterizes stationary regime of batches motion. This regime describes intensionally such a situation when after the long-duration interval of time motorcades following one another move along the motorway and don't disturb their structure. Let's take notes that such distribution doesn't depend on the entry conditions and obeys the normalization condition: $\sum\limits_{k=1}^{N} Q(k) = 1$. Expressing $Q(k)$ in terms of $Q(1)$ and substituting derived expressions in the normalization condition we can ascertain that ergodic probability distribution depends on only three essential parameters: $\nu_1 = \frac{\lambda_0}{\mu_1}, \nu_2 = \frac{\lambda_0}{\mu_2}, \nu_3 = \frac{\lambda_0}{\mu_3}$. Finally we derive the distribution of batches motion in the stationary regime:

$$\begin{gathered} Q(1) = (1 + \nu_1 + \nu_1\nu_2 \frac{\nu_3^{N-2} - 1}{\nu_3 - 1})^{-1}, Q(2) = \nu_1 (1 + \nu_1 + \nu_1\nu_2 \frac{\nu_3^{N-2} - 1}{\nu_3 - 1})^{-1}, \\ Q(3) = \nu_1\nu_2 (1 + \nu_1 + \nu_1\nu_2 \frac{\nu_3^{N-2} - 1}{\nu_3 - 1})^{-1}, \\ Q(k) = \nu_1\nu_2\nu_3^{k-3} (1 + \nu_1 + \nu_1\nu_2 \frac{\nu_3^{N-2} - 1}{\nu_3 - 1})^{-1}, k = 4, 5, \ldots, N. \end{gathered}$$

In particular case when $N = 3$ and $\mu_2 = \mu_3$ system (2) takes on form:

$$\begin{gathered} dQ(t,1)/dt = -\lambda_0 Q(t,1) + \mu_1 Q(t,2), \\ dQ(t,2)/dt = \lambda_0 Q(t,1) - (\lambda_0 + \mu_1) Q(t,2) + \mu_2 Q(t,3), \\ dQ(t,3)/dt = \lambda_0 Q(t,2) - \mu_2 Q(t,3). \end{gathered}$$

Explicit solution of this system can be found in different ways. For example, one of them is adduced in monograph [3]. We should notice that explicit solution is a linear combination of three independent particular solutions of this system. The coefficients

of the given combination should be derived from entry conditions. If we proceed to the limit $t \to \infty$ in the explicit expression of system (2) solution, the following equalities will obtain

$$Q(1) = (1 + \nu_1 + \nu_1\nu_2)^{-1}, Q(2) = \nu_1(1 + \nu_1 + \nu_1\nu_2)^{-1},$$
$$Q(3) = \nu_1\nu_2(1 + \nu_1 + \nu_1\nu_2)^{-1}.$$

We would have had the same results if we had proceeded to the limit in the very system as it has been shown above for the whatever N.

3. PROPERTIES OF THE FLOW

We have studied the single traffic batch so now let's return to the whole traffic flow. Watching the light traffic flow in practice we can observe that density of the slow vehicles exceeds density of the fast vehicles appreciably. At the same time the slow vehicles move independently in the stationary mode. Also we consider that on average distance between consecutive slow vehicles is lot more than maximal length of road section occupied by an average number of vehicles in non-homogeneous traffic batch. That's why it is possible to consider that flow of the slow vehicles is the Poisson flow with a parmeter λ and all the vehicles in the batch cross a transverse line in the stationary regime simultaneously. Such a flow is acceptedly called non-ordinary Poisson flow. Let's introduce a random variable $\eta(t) = \eta(\omega; t)$ which counts a number of all vehicles crossed a stop–line in the time interval $[0, t)$ and let's designate $P_k(t) = \mathbf{P}(\eta(t) = k)$. We will think traffic flow is non-ordinary Poisson flow in case when in every calling instant one arrival comes with a probability $p = (1+\nu_1+\nu_1\nu_2)^{-1}$, two arrivals – with a probability $q = \nu_1(1 + \nu_1 + \nu_1\nu_2)^{-1}$ and three – with a probability $s = \nu_1\nu_2(1 + \nu_1 + \nu_1\nu_2)^{-1}$.

Theorem 1. *For the probability generating function*

$$\Psi(t, z) = \sum_{k=0}^{\infty} P_k(t) z^k$$

of random variable $\eta(\omega; t)$ distribution the following equality takes place

$$\Psi(t, z) = e^{-\lambda t} \sum_{k=0}^{\infty} z^k \sum_{i=0}^{[\frac{k}{2}]} \sum_{j=0}^{[\frac{k-2i}{3}]} \binom{k-i-2j}{i, j, k-2i-3j} p^{k-2i-3j} q^i s^j \frac{(\lambda t)^{k-i-2j}}{(k-i-2j)!},$$

where $[x]$ means integer part of x.

Theorem 2. *For the one-dimensional distributions $P_k(t)$ the following equalities are true*

$$P_0(t) = e^{-\lambda t}, P_1(t) = \lambda t p e^{-\lambda t},$$

$$P_k(t) = e^{-\lambda t} \sum_{i=0}^{[\frac{k}{2}]} \sum_{j=0}^{[\frac{k-2i}{3}]} \binom{k-i-2j}{i, j, k-2i-3j} p^{k-2i-3j} q^i s^j \frac{(\lambda t)^{k-i-2j}}{(k-i-2j)!}, k = 2, 3, \ldots$$

Lemma 1. *The sum of m independent non-ordinary Poisson flows with parameters λ_j, p_j and q_j (where $j = 1, 2, \ldots, m$) is non-ordinary Poisson flow with parameters*

$$\lambda = \sum_{j=0}^{m} \lambda_j,\ p = (\sum_{j=0}^{m} \lambda_j p_j)/(\sum_{j=0}^{m} \lambda_j),\ q = (\sum_{j=0}^{m} \lambda_j q_j)(\sum_{j=0}^{m} \lambda_j).$$

Lemma 2. *For the expectation and dispersion of random variable $\eta(\omega; t)$ the following qualities are true*

$$\mathbf{M}\eta(\omega; t) = \lambda t(1 + q + 2s),\ \mathbf{D}\eta(\omega; t) = \lambda t(1 + 3q + 8s).$$

4. CONCLUSION

Problem concerned construction and investigation of mathematical model of traffic flow spatial disposition on the crossroads-free motorway is solved. The nonlocal description method for the unconditioned flow is used. So not the probabilistic properties of every vehicle random disposition on the motorway but properties of some group of non-homogeneous vehicles random disposition are defined. For this purpose the easy mechanism of small traffic batches formation in case of bad weather and bad road conditions is proposed. It is proved that temporal characteristic of traffic flow stationary motion can be locally described as the non-ordinary Poisson flow with a limited number of the arrivals received in every calling instant. The superposition property of a finite number of such flows is proved. Easy formulas for basic numerical characteristics of such flows are found.

REFERENCES

1. *Haight F. A.* Mathematical theories of traffic flow. New York London: Academic press, 1963.
2. *Fedotkin M. A.* Nonlocal way to define the controllable stochastic processes // Mathematical questions of the cybernetics, 1998. № 7. P. 332–344.
3. *Klimov G.P.* Stochastic service systems. M.: Science, 1966.

АЛГОРИТМ ОЦЕНКИ ПАРАМЕТРОВ АСИНХРОННОГО ПОТОКА СОБЫТИЙ С КОНЕЧНЫМ ЧИСЛОМ СОСТОЯНИЙ

А. Горцев, В. Зуевич*

Томский государственный университет
Томск, Россия
* zuevichv@ya.ru

Получены оптимальные оценки параметров асинхронного дважды стохастического потока с конечным числом состояний. Оценки оптимальны в смысле минимума среднеквадратического отклонения от истинных значений параметров потока. Оценивание производится на основе наблюдений за моментами наступления событий потока. Приведен алгоритм оценивания.

Ключевые слова: асинхронный дважды стохастический поток, MMPP-поток, оптимальная оценка параметров, алгоритм оценки параметров, цифровые сети интегрального обслуживания, ISDN.

1. ВВЕДЕНИЕ

Интенсивное развитие информационных технологий определило важную сферу приложений теории массового обслуживания – проектирование и создание компьютерных сетей связи, спутниковых сетей связи, телекоммуникационных сетей, информационно-вычислительных сетей и т.п., которых можно объединить одним термином – цифровые сети интегрального обслуживания (ЦСИО, ISDN). Возникла необходимость в разработке новых математических моделей потоков событий, адекватно описывающих реальные информационные потоки, функционирующие в ЦСИО. Одними из первых работ в этом направлении были [1, 2, 3]. Подчеркнем, что на практике параметры, характеризующие поток событий, частично либо полностью неизвестны, а также могут изменяться с течением времени случайным образом, что приводит к рассмотрению дважды стохастических потоков событий. Поскольку функционирование системы обслуживания непосредственно зависит от параметров входящего потока, важной задачей является оценка в произвольный момент времени параметров входящего потока по наблюдениям за этим потоком. Исследования по оценке параметров дважды стохастических потоков были проведены, например, в работах [4, 5, 6].

В настоящей статье решена задача оценки парамтеров асинхронного дважды стохастического потока событий с конечным числом состояний.

2. ПОСТАНОВКА ЗАДАЧИ

В статье рассматривается асинхронный дважды стохастический поток событий с произвольным конечным числом состояний [7] (далее асинхронный поток либо просто поток). Интенсивность потока является кусочно-постоянным случайным процессом $\lambda(t)$ с n состояниями: $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$. Процесс (поток) в момент времени t находится в i-м состоянии, если $\lambda(t) = \lambda_i$ $(i = \overline{1,n})$. В течение времени пребывания в i-м состоянии поток ведет себя как пуассоновский с интенсивностью λ_i $(i = \overline{1,n})$. Длительность пребывания в i-м состоянии есть экспоненциально распределенная случайная величина с функцией распределения $F_i(\tau) = 1 - e^{\alpha_{ii}\tau}$, где $\alpha_{ii} = -\sum_{j=1, j\neq i}^{n} \alpha_{ij}$ $(i = \overline{1,n})$; $\alpha_{ij} > 0$ $(i,j = \overline{1,n}, i \neq j)$ – интенсивность перехода процесса $\lambda(t)$ из состояния i в состояние j, т.е. величины α_{ij} образуют матрицу интенсивностей (матрицу инфинитезимальных коэффициентов) переходов между состояниями $\|\alpha_{ij}\|_1^n$. В сделанных предпосылках $\lambda(t)$ — транзитивный марковский процесс [7].

Значения параметров потока λ_i, α_{ij} $(i,j = \overline{1,n}, i \neq j)$ неизвестны. Процесс $\lambda(t)$ является принципиально ненаблюдаемым. Предполагается, что известно только число состояний n и наблюдению доступны моменты наступления событий потока $t_1, t_2, \ldots$. Необходимо по наблюдениям $t_1, t_2, \ldots$ оценить параметры потока λ_i, α_{ij} $(i,j = \overline{1,n}, i \neq j)$ в момент окончания наблюдения за потоком.

Рассматривается стационарный (установившийся) режим функционирования наблюдаемого потока событий, поэтому переходными процессами на интервале наблюдения (t_0, t), где t_0 – начало наблюдений, t – окончание наблюдений (момент вынесения решения), пренебрегаем. Без потери общности можно положить $t_0 = 0$.

Обозначим $\theta = (\lambda_1, \ldots, \lambda_n, \alpha_{ij}; i,j = \overline{1,n}, i \neq j)$ – вектор неизвестных параметров потока, $\hat{\theta}(t)$ – вектор соответствующих оценок параметров в момент времени t. θ_k и $\hat{\theta}_k(t)$ – k-е компоненты вектора параметров и вектора оценок соответственно. Обозначим N – размерность векторов θ и $\hat{\theta}(t)$, $N = n^2$. Обозначим $p(\theta|t) = p(\theta|t_1, t_2, \ldots, t_m)$ – апостериорная плотность распределения вероятностей вектора параметров θ в момент времени t при условии, что в моменты времени $t_1, t_2, \ldots, t_m$ $(0 < t_1 < t_2 < \ldots < t_m < t)$ наблюдались события потока. Обозначим $\Theta = \{\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0,\ \alpha_{ij} > 0;\ i,j = \overline{1,n}, i \neq j\}$ – область значений вектора параметров θ. Будем использовать оценку $\hat{\theta}$, оптимальную в смысле минимума среднеквадратического отклонения от истинного значения вектора θ:

$$\hat{\theta}_k(t) = \int\limits_{\Theta} \theta_k p(\theta|t)\, d\theta \ \ (k = \overline{1,N}),$$

где $d\theta = d\theta_1 d\theta_2 \ldots d\theta_N$, или в векторной форме

$$\hat{\theta}(t) = \int\limits_{\Theta} \theta p(\theta|t)\, d\theta. \tag{1}$$

Выражение (1) дает оптимальную оценку параметров потока в виде апостериорного среднего. Для нахождения оценки параметров по формуле (1) необходимо найти выражение для апостериорной плотности $p(\theta|t)$.

3. ПОЛУЧЕНИЕ ЯВНОГО ВИДА АПОСТЕРИОРНОЙ ПЛОТНОСТИ ВЕРОЯТНОСТЕЙ

Для получения рекуррентной формулы воспользуемся известной методикой: рассмотрим дискретные наблюдения за потоком через равные достаточно малые промежутки времени длительности Δt, а затем устремим Δt к нулю. Пусть наблюдения за потоком начинаются в момент времени $t = 0$, и время t изменяется дискретно с конечным шагом Δt: $t = k\Delta t$ $(k = 0, 1, \ldots)$. Обозначим $r(k\Delta t)$ – число событий, наблюденных на полуинтервале времени $[(k–1)\Delta t, k\Delta t)$, $r(k\Delta t) = 0, 1, \ldots,$ $(k = 0, 1, \ldots)$. На полуинтервале $[–\Delta t, 0)$ наблюдений не производится, поэтому $r(0)$ можем положить произвольным, например положим $r(0) = 0$. Обозначим $\vec{r}(m\Delta t) = (r(0), r(\Delta t), \ldots, r(m\Delta t))$ – последовательность значений количества наблюденных событий на временных полуинтервалах $[(k–1)\Delta t, k\Delta t)$ $(k = \overline{0, m})$. Рассмотрим момент времени t такой, что $t = m\Delta t, t + \Delta t = (m + 1)\Delta t$. Тогда имеем $r(m\Delta t) = r(t)$, $r((m + 1)\Delta t) = r(t + \Delta t)$, $\vec{r}(m\Delta t) = \vec{r}(t)$, $\vec{r}((m + 1)\Delta t) = \vec{r}(t + \Delta t)$.

Рассмотрим $p(\theta|\vec{r}(m\Delta t)) = p(\theta|\vec{r}(t)) = p(\theta|t)$ – апостериорную плотность вероятностей вектора параметров θ в момент времени t при условии, что на полуинтервалах времени $[(k–1)\Delta t, k\Delta t)$ $(k = \overline{0, m})$ наблюдалось $r(k\Delta t)$ событий потока соответственно. Также рассмотрим $p(\theta|t + \Delta t)$ – апостериорную плотность в момент времени $t + \Delta t$.

Теорема 1. *Апостериорная плотность $p(\theta|t+\Delta t)$ определяется рекуррентной формулой*

$$p(\theta|t + \Delta t) = p(\theta|t)\frac{\sum_{j=1}^{n} \omega(\lambda_j|t)\frac{(\lambda_j\Delta t)^{r(t+\Delta t)}}{r(t+\Delta t)!}e^{-\lambda_j\Delta t}}{\int\limits_{\Theta} p(\theta|t)\sum_{j=1}^{n} \omega(\lambda_j|t)\frac{(\lambda_j\Delta t)^{r(t+\Delta t)}}{r(t+\Delta t)!}e^{-\lambda_j\Delta t}\, d\theta}, \tag{2}$$

где $\omega(\lambda_j|t)$ $(j = \overline{1, n})$ – апостериорная вероятность того, что поток в момент времени t $(t > t_0)$ находится в j-м состоянии; $\omega(\lambda_j|t)$ $(j = \overline{1, n})$ получены в явном виде в работе [7].

Доказательство. Используя формулу для условной вероятности, находим

$$p(\theta|\vec{r}(t + \Delta t)) = \frac{p(\theta, r(t + \Delta t), \vec{r}(t))}{p(\vec{r}(t + \Delta t))} = \frac{p(r(t + \Delta t)|\theta, \vec{r}(t))p(\theta, |\vec{r}(t))p(\vec{r}(t))}{p(\vec{r}(t + \Delta t))},$$

или, используя в последней дроби формулу полной вероятности для условной вероятности, получаем

$$p(\theta|\vec{r}(t+\Delta t)) = p(\theta|\vec{r}(t))\frac{p(\vec{r}(t))}{p(\vec{r}(t + \Delta t))}\sum_{j=1}^{n} p(r(t+\Delta t)|\vec{r}(t), \theta, \lambda(t) = \lambda_j)p(\lambda(t) = \lambda_j|\vec{r}(t), \theta). \tag{3}$$

Рассмотрим в (3) произведение $p(r(t + \Delta t)|\vec{r}(t), \theta, \lambda(t) = \lambda_j)p(\lambda(t) = \lambda_j|\vec{r}(t), \theta)$. Во-первых, имеем $p(\lambda(t) = \lambda_j|\vec{r}(t), \theta) = \omega(\lambda_j|t)$ $(j = \overline{1, n})$, где $\omega(\lambda_j|t)$ $(j = \overline{1, n})$ – апостериорная вероятность того, что поток в момент времени t находится в j-м

состоянии, определенная в [7]. Во-вторых, количество наблюденных на полуинтервале $[t, t+\Delta t)$ событий не зависит от последовательности $\vec{r}(t)$ наблюденных до момента t событий, а зависит только от состояния потока в момент времени t, т.е. от значения интенсивности наступления событий потока $\lambda(t) = \lambda_j \ \ (j = \overline{1,n})$ в момент времени t. Поэтому имеем $p(r(t+\Delta t)|\vec{r}(t), \theta, \lambda(t) = \lambda_j) = p(r(t+\Delta t)|\theta, \lambda(t) = \lambda_j)$. А поскольку поток в любом состоянии ведет себя как пуассоновский, получаем $p(r(t+\Delta t)|\theta, \lambda(t) = \lambda_j) = \frac{(\lambda_j \Delta t)^{r(t+\Delta t)}}{r(t+\Delta t)!} e^{-\lambda_j \Delta t}$. Таким образом, получаем (3) в виде

$$p(\theta|\vec{r}(t+\Delta t)) = p(\theta|\vec{r}(t)) \frac{p(\vec{r}(t))}{p(\vec{r}(t+\Delta t))} \sum_{j=1}^{n} \omega(\lambda_j|t) \frac{(\lambda_j \Delta t)^{r(t+\Delta t)}}{r(t+\Delta t)!} e^{-\lambda_j \Delta t}.$$

Находя в последнем выражении множитель $p(\vec{r}(t))/p(\vec{r}(t+\Delta t))$ из условия нормировки $\int\limits_{\Theta} p(\theta|\vec{r}(t+\Delta t))\, d\theta = 1$, приходим к (2). □

Рассмотрим случай, когда на полуинтервале $[t, t+\Delta t)$ нет событий, т.е. $r(t+\Delta t) = 0$. Это означает, что полуинтервал $[t, t+\Delta t)$ находится на временной оси между моментами наступления соседних событий, скажем, между моментами t_k и t_{k+1} $(k = 1, 2, \ldots)$ либо между моментом начала наблюдений за потоком t_0 и моментом наступления первого события t_1. Для дальнейших выкладок введем обозначение $s(t, \theta) = \sum_{j=1}^{n} \lambda_j \omega(\lambda_j|t)$.

Лемма 1. *Апостериорная плотность $p(\theta|t)$ между моментами наступления событий удовлетворяет интегро-дифференциальному уравнению*

$$\frac{dp(\theta|t)}{dt} = -p(\theta|t) \left[s(t,\theta) - \int\limits_{\Theta} s(t,\theta) p(\theta|t)\, d\theta \right] \quad (t_k < t < t_{k+1}, \ \ k = 0, 1, \ldots). \qquad (4)$$

Доказательство. Поскольку $r(t+\Delta t) = 0$, получаем (2) в виде

$$p(\theta|t+\Delta t) = p(\theta|t) \frac{\sum_{j=1}^{n} \omega(\lambda_j|t) e^{-\lambda_j \Delta t}}{\int\limits_{\Theta} p(\theta|t) \sum_{j=1}^{n} \omega(\lambda_j|t) e^{-\lambda_j \Delta t}\, d\theta},$$

или, раскладывая экспоненты в числителе и знаменателе в ряд с точностью до $o(\Delta t)$, получаем

$$p(\theta|t+\Delta t) = [p(\theta|t) - s(t,\theta) p(\theta|t) \Delta t + o(\Delta t)] \left[1 - \Delta t \int\limits_{\Theta} s(t,\theta) p(\theta|t)\, d\theta + o(\Delta t) \right]^{-1},$$

или, раскладывая второй сомножитель в ряд, находим

$$p(\theta|t+\Delta t) = p(\theta|t) - \Delta t s(t,\theta) p(\theta|t) + \Delta t p(\theta|t) \int\limits_{\Theta} s(t,\theta) p(\theta|t))\, d\theta + o(\Delta t).$$

Перенося в последнем равенстве $p(\theta|t)$ влево, деля на Δt и устремляя Δt к нулю, получаем (4). □

Асинхронный поток обладает свойством ординарности (поскольку в каждом состоянии ведет себя как пуассоновский), поэтому вероятность наступления на полуинтервале $[t, t+\Delta t)$ двух и более событий равна $o(\Delta t)$. Рассмотрим случай, когда на полуинтервале $[t, t+\Delta t)$ наступает одно событие потока, скажем, в момент времени t_k $(t < t_k < t+\Delta t)$. При этом $r(t+\Delta t) = 1$.

Лемма 2. *Апостериорная плотность $p(\theta|t)$ в моменты наступления событий пересчитывается по формуле*

$$p(\theta|t_k+0) = \frac{p(\theta|t_k-0)s(t_k-0,\theta)}{\int\limits_{\Theta} p(\theta|t_k-0)s(t_k-0,\theta)\,d\theta} \quad (k = 1, 2, \ldots). \tag{5}$$

Доказательство. Поскольку $r(t+\Delta t) = 1$, получаем (2) в виде

$$p(\theta|t+\Delta t) = p(\theta|t)\frac{\Delta t \sum_{j=1}^{n} \omega(\lambda_j|t)\lambda_j e^{-\lambda_j \Delta t}}{\Delta t \int\limits_{\Theta} p(\theta|t) \sum_{j=1}^{n} \omega(\lambda_j|t)\lambda_j e^{-\lambda_j \Delta t}\,d\theta},$$

или, раскладывая экспоненты в числителе и знаменателе в ряд с точностью до $o(\Delta t)$, и вводя величины $\Delta t'$ и $\Delta t''$ такие, что $t = t_k - \Delta t', t+\Delta t = t_k + \Delta t''$, находим

$$p(\theta|t_k+\Delta t'') = p(\theta|t)\frac{\Delta t p(\theta|t_k-\Delta t')s(t_k-\Delta t',\theta) + o(\Delta t)}{\Delta t \int\limits_{\Theta} p(\theta|t_k-\Delta t')s(t_k-\Delta t',\theta)\,d\theta + o(\Delta t)}.$$

Поделим числитель и знаменатель последней дроби на Δt, после чего устремим Δt к нулю (при этом $t = t_k - \Delta t'$ стремится к t_k слева, $t+\Delta t = t_k + \Delta t''$ стремится к t_k справа). После предельного перехода получаем (5). □

Замечание к лемме 2. В момент времени t_0 начала наблюдений за потоком апостериорная плотность $p(\theta|t_0)$ задается исходя из априорных данных о параметрах асинхронного потока. Если таких данных нет, можно задать плотность $p(\theta|t_0)$ как произведение N плотностей равномерно распределенных случайных величин, каждая из которых распределена в некотором интервале допустимых значений для соответствующего параметра потока.

Теорема 2. *Поведение апостериорной плотности $p(\theta|t)$ на временной оси определяется интегро-дифференциальным уравнением* (4) *и формулой пересчета плотности вероятностей* (5), *в которых $t_k \le t < t_{k+1}$, $p(\theta|t_k) = p(\theta|t_k+0)$, $p(\theta|t_{k+1}) = p(\theta|t_{k+1}-0)$ $(k = 0, 1, \ldots)$. В момент времени t_0 апостериорная плотность $p(\theta|t_0)$ задается согласно замечанию к лемме 2.*

Доказательство. Доказательство следует из лемм 1 и 2 путем синхронизации формул (4) и (5). □

Теорема 3. *Апостериорная плотность вероятностей $p(\theta|t)$ вектора параметров θ на полуинтервале времени $[t_k, t_{k+1})$ $(k = 0, 1, \ldots)$ определяется формулой:*

$$p(\theta|t) = \frac{p(\theta|t_k)\mathbf{e}^{\left[-\int\limits_{t_k}^{t} s(\tau,\theta)\,d\tau\right]}}{\int\limits_{\Theta} p(\theta|t_k)\mathbf{e}^{\left[-\int\limits_{t_k}^{t} s(\tau,\theta)\,d\tau\right]}\,d\theta} \quad (t_k < t < t_{k+1}, k = 0, 1, \ldots), \quad (6)$$

где $p(\theta|t_k) = p(\theta|t_k + 0)$ вычисляется в момент наступления события t_k $(k = 1, 2, \ldots)$ по формуле (5), а в момент t_0 плотность $p(\theta|t_0)$ задается согласно замечанию к лемме 2.

Доказательство. Преобразуем интегро-дифференциальное уравнение (4) к следующему виду:

$$\frac{dp(\theta|t)}{p(\theta|t)} = -\left[s(t,\theta) - \int\limits_{\Theta} s(t,\theta)p(\theta|t))\,d\theta\right]dt.$$

Интегрируя последнее интегро-дифференциальное уравнение в пределах от t_k до t, получаем

$$\ln\left[\frac{p(\theta|t)}{p(\theta|t_k)}\right] = -\int\limits_{t_k}^{t}\left[s(\tau,\theta) - \int\limits_{\Theta} s(\tau,\theta)p(\theta|\tau)\,d\theta\right]d\tau,$$

или, проделывая необходимые преобразования, получаем

$$p(\theta|t) = p(\theta|t_k)\mathbf{e}^{\left[-\int\limits_{t_k}^{t} s(\tau,\theta)\,d\tau\right]}\mathbf{e}^{\left[\int\limits_{t_k}^{t}\int\limits_{\Theta} s(\tau,\theta)p(\theta|\tau))\,d\theta\,d\tau\right]}.$$

Найдя в полученном выражении последний сомножитель из условия нормировки $\int\limits_{\Theta} p(\theta|t)\,d\theta = 1$, приходим к (6). □

Подставляя (6) в (1), получаем явный вид оценок параметров потока для произвольного момента времени наблюдения за потоком. При этом в момент времени t_{k+1} $(k = 0, 1, \ldots)$ имеем

$$\hat{\theta}(t_{k+1}) = \hat{\theta}(t_{k+1} + 0) = \int\limits_{\Theta} \theta p(\theta|t_{k+1} + 0)\,d\theta, \quad (7)$$

где $p(\theta|t_k + 0)$ вычисляется по формуле (5).

4. АЛГОРИТМ ОЦЕНКИ ПАРАМЕТРОВ ПОТОКА

Полученные в настоящей статье формулы позволяют сформулировать алгоритм расчета оценок параметров асинхронного дважды стохастического потока событий с конечным числом состояний в произвольный момент времени t наблюдения за потоком:

1) в момент времени $t_0 = 0$ задается $p(\theta|t_0)$, исходя из априорных данных о потоке, либо как произведение N плотностей равномерно распределенных случайных величин, каждая из которых распределена в некотором интервале допустимых значений для соответствующего параметра потока; также задаются начальные значения апостериорных вероятностей $\omega(\lambda_j|t_0)$, $(i = \overline{1,n})$, способ задания апостериорных вероятностей описан в [7];
2) по формуле (1) в момент времени $t_0 = 0$ рассчитываются оценки параметров потока $\hat{\theta}(t_0)$;
3) для $k = 0$ в любой момент времени t $(0 < t < t_{k+1})$, где t_{k+1} – момент наблюдения $(k+1)$-го события потока, по формуле (6) рассчитывается апостериорная плотность $p(\theta|t)$ и параллельно с этим по формуле (1) рассчитываются оценки параметров потока $\hat{\theta}(t)$; при этом вероятности $\omega(\lambda_j|t)$ $(j = \overline{1,n})$, входящие в (6), вычисляются согласно [7];
4) для $k = 0$ в момент времени t_{k+1} по формуле (5) рассчитывается апостериорная плотность $p(\theta|t_{k+1}+0)$ и по формуле (7) находятся оценки $\hat{\theta}(t_{k+1})$; при этом $p(\theta|t_{k+1}-0)$ вычисляется по формуле (6) в момент времени $t = t_{k+1}$, вероятности $\omega(\lambda_j|t)$ $(j = \overline{1,n})$, входящие в (6), в момент $t = t_{k+1}$ вычисляются согласно [7]; значение $p(\theta|t_{k+1}+0)$ является начальным для расчета плотности $p(\theta|t)$ на следующем шаге алгоритма;
5) шаги 3 – 4 алгоритма повторяются для $k = 1, 2, \ldots$.

5. ЗАКЛЮЧЕНИЕ

В настоящей статье получены оптимальные оценки параметров асинхронного дважды стохастического потока с конечным числом состояний [7]. Оценки оптимальны в смысле минимума среднеквадратического отклонения от истинных значений параметров потока. Приводится алгоритм оценки параметров асинхронного потока событий. Сформулированный алгоритм может быть реализован программно и применяться для оценки параметров реальных потоков. Оценки параметров реальных потоков используются при создании адекватных математических моделей реально существующих и проектируемых ЦСИО.

ЛИТЕРАТУРА

1. *Башарин Г. П., Кокотушкин В. А., Наумов В. А.* О методе эквивалентных замен расчета фрагментов сетей связи. Ч.1 // Изв. АН СССР. Техн. кибернетика, 1979. № 6. С. 92–99.

2. *Башарин Г. П., Кокотушкин В. А., Наумов В. А.* О методе эквивалентных замен расчета фрагментов сетей связи. Ч.2 // Изв. АН СССР. Техн. кибернетика, 1980. № 1. С. 55–61.
3. *Neuts M. F.* A versatile Markov point process // J. Appl. Probab., 1979. V. 16. P. 764–779.
4. *Горцев А. М., Нежельская Л. А.* Оценка параметров синхронного МС-потока событий // Сети связи и сети ЭВМ (анализ и применение): тез. докл. Восьмой Белорусской зимней школы-семинара по теории массового обслуживания, Минск: Изд-во БГУ, 1992. С. 33.
5. *Горцев А. М., Шмырин И. С.* Оптимальная оценка параметров дважды стохастического пуассоновского потока событий при наличии ошибок в измерениях моментов наступления событий // Изв. вузов. Физика, 1999. № 4. С. 19–27.
6. *Бушланов И. В., Горцев А. М., Нежельская Л. А.* Оценка параметров синхронного дважды стохастического потока событий // АиТ, 2008. № 9. С. 76–93.
7. *Горцев А. М., Зуевич В. Л.* Оптимальная оценка состояний асинхронного дважды стохастического потока событий с произвольным числом состояний // Вестник ТГУ. Управление, вычислительная техника и информатика, 2010. № 2(11). С. 44–65.

ОПТИМАЛЬНАЯ ОЦЕНКА СОСТОЯНИЙ ОБОБЩЕННОГО ПОЛУСИНХРОННОГО ПОТОКА ПРИ НЕПРОДЛЕВАЮЩЕМСЯ МЕРТВОМ ВРЕМЕНИ

А. Калягин, Л. Нежельская

Томский Государственный Университет
Томск, Россия
redall@inbox.ru

Рассматривается задача оптимальной оценки состояний полусинхронного дважды стохастического потока с инициированием дополнительных событий (обобщенный полусинхронный поток событий) с двумя состояниями. Условия наблюдения за потоком таковы, что каждое событие порождает период мертвого времени, в течении которого другие события потока недоступны наблюдению и не вызывают продления его периода (непродлевающееся мертвое время). Находится явный вид апостериорных вероятностей состояний потока. Решение о состоянии потока выносится по критерию максимума апостериорной вероятности. Приводятся численные результаты, полученные с использованием расчетных формул и имитационного моделирования.

Ключевые слова: обобщенный полусинхронный поток событий, состояние потока, апостериорная вероятность состояния, оценка состояния, мёртвое время.

1. ВВЕДЕНИЕ

В настоящей статье проводится дальнейшее исследование обобщенного полусинхронного потока событий, начатое в работе [1]. Обобщенный полусинхронный дважды стохастический поток событий (далее обобщенный полусинхронный поток либо просто поток) является одной из адекватных математических моделей информационных потоков заявок, функционирующих в цифровых сетях интегрального обслуживания [2]. Достаточно обширная литература по исследованию подобных потоков событий (асинхронных, синхронных и полусинхронных) приведена в [1, 3, 4].

Подавляющее число авторов рассматривает математические модели потоков событий, когда все события доступны наблюдению. Однако на практике возникают ситуации, когда наступившее событие может повлечь за собой ненаблюдаемость последующих событий. Одним из искажающих факторов при оценке состояний и параметров потока событий выступает мертвое время регистрирующих приборов [5], которое порождается зарегистрированным событием. Другие же события,

наступившие в течение периода мертвого времени, недоступны наблюдению (теряются). Можно считать, что этот период продолжается некоторое фиксированное время T. Все устройства регистрации делятся на две группы. Первую группу составляют устройства с непродлевающимся мертвым временем, вторую - устройства с продлевающимся мертвым временем. Отметим, что одними из первых работ по оценке параметров в случайных потоках событий, функционирующих в условиях мертвого времени, являются работы [6, 7]. В работе [1] решена задача об оптимальной оценке состояний обобщенного полусинхронного потока событий в условиях отсутствия мертвого времени. В настоящей статье, являющейся непосредственным развитием работы [1], решается задача об оптимальной оценке состояний обобщенного полусинхронного потока событий в условиях его неполной наблюдаемости (при непродлевающемся мертвом времени). Предлагается алгоритм оптимальной оценки состояний, когда решение о состоянии обобщенного полусинхронного потока выносится по критерию максимума апостериорной вероятности, представляющей наиболее полную характеристику состояния потока, которую можно получить, располагая только выборкой наблюдений, и обеспечивающего минимум полной (безусловной) вероятности ошибки вынесения решения [8].

2. ПОСТАНОВКА ЗАДАЧИ

Рассматривается обобщенный полусинхронный поток событий, интенсивность которого есть кусочно-постоянный случайный процесс $\lambda(t)$ с двумя состояниями λ_1 и λ_2 ($\lambda_1 > \lambda_2$). В течение временного интервала, когда $\lambda(t) = \lambda_i$, имеет место пуассоновский поток событий с интенсивностью λ_i , $i = 1, 2$. Переход из первого состояния процесса $\lambda(t)$ во второе возможен только в момент наступления события, при этом переход осуществляется с вероятностью $p(0 < p \leq 1)$; с вероятностью $1 - p$ процесс $\lambda(t)$ остается в первом состоянии. Тогда длительность пребывания процесса $\lambda(t)$ в первом состоянии есть случайная величина с экспоненциальной функцией распределения $F_1(\tau) = 1 - e^{-p\lambda_1\tau}$. Переход из второго состояния процесса $\lambda(t)$ в первое состояние может осуществляться в произвольный момент времени. При этом длительность пребывания процесса $\lambda(t)$ во втором состоянии распределена по экспоненциальному закону: $F_2(\tau) = 1 - e^{-\alpha\tau}$. При переходе процесса $\lambda(t)$ из второго состояния в первое инициируется с вероятностью δ $(0 \leq \delta \leq 1)$ дополнительное событие в первом состоянии (т.е. сначала осуществляется переход, а затем инициируется дополнительное событие). Очевидно, что в сделанных предпосылках $\lambda(t)$ - марковский процесс. После каждого зарегистрированного события в момент времени t_i наступает время фиксированной длительности T (мертвое время), в течение которого другие события исходного обобщенного полусинхронного потока событий недоступны наблюдению. События, наступившие в течение мертвого времени, не вызывают продления его периода (непродлевающееся мертвое время). По окончании мертвого времени первое наступившее событие снова создает период мертвого времени длительности T и т.д.

Так как процесс $\lambda(t)$ и типы событий (события пуассоновских потоков и дополнительные события) являются принципиально ненаблюдаемыми, а наблюдаемыми являются только временные моменты наступления событий наблюдаемого потока $t_1, t_2, \ldots$, то необходимо по этим наблюдениям оценить состояние процесса (потока) $\lambda(t)$ в момент окончания наблюдений.

Рассматривается установившийся (стационарный) режим функционирования наблюдаемого потока событий, поэтому переходными процессами на интервале наблюдения (t_0, t), где t_0 - начало наблюдений, t - окончание наблюдений (момент вынесения решения), пренебрегаем. Тогда без потери общности можно положить $t_0 = 0$. Для вынесения решения о состоянии ненаблюдаемого процесса $\lambda(t)$ в момент времени t необходимо определить апостериорные вероятности $w(\lambda_i \mid t) = w(\lambda_i \mid t_1, \ldots, t_m, t)$, $i = 1, 2$, того, что в момент времени t значение процесса $\lambda(t) = \lambda_i$ (m - количество наблюденных событий за время t), при этом $w(\lambda_1 \mid t) + w(\lambda_2 \mid t) = 1$. Решение о состоянии процесса $\lambda(t)$ выносится путем сравнения апостериорных вероятностей: если $w(\lambda_j \mid t) \geq w(\lambda_i \mid t)$, $i, j = 1, 2$, $i \neq j$, то оценка состояния процесса есть $\lambda(t) = \lambda_j$.

3. АЛГОРИТМ ОПТИМАЛЬНОЙ ОЦЕНКИ СОСТОЯНИЙ

Момент вынесения решения t будет принадлежать интервалу (t_i, t_{i+1}), $i = 0, 1, 2, \ldots$, между двумя соседними событиями наблюдаемого потока. Рассмотрим интервал времени (t_i, t_{i+1}), значение длительности которого есть $\tau_i = t_{i+1} - t_i$, $i = 0, 1, \ldots$. Так как моменты наступления событий в наблюдаемом потоке случайны, то длительность интервала (t_i, t_{i+1}) - случайная величина. С другой стороны, так как наблюдаемое в момент времени t_i событие порождает период мертвого времени длительности T, то $\tau_i = T + \eta_i$, где η_i - значение длительности интервала между моментом окончания периода мертвого времени и моментом t_{i+1}, т.е. интервал (t_i, t_{i+1}) разбивается на два смежных интервала: первый - $(t_i, t_i + T)$, второй - $(t_i + T, t_{i+1})$. Отметим одно важное обстоятельство: так как после каждого события в наблюдаемом потоке реализуется период мертвого времени длительности T, в течение которого последующие события обобщенного полусинхронного потока недоступны наблюдению (поток отсутствует), то условия нахождения апостериорной вероятности $w(\lambda_1 \mid t)$ на интервале $(t_i, t_i + T)$ длительности T и интервале $(t_i + T, t_{i+1})$, значение длительности которого есть η_i, принципиально разные. Кроме того, для нахождения апостериорной вероятности $w(\lambda_1 \mid t)$ необходимо точно знать значение длительности T мертвого времени либо, по крайней мере, предварительно осуществить её оценку $\widehat{T}$. В противном случае отсутствие информации о значении длительности T мертвого времени делает попытку строгого нахождения апостериорной вероятности $w(\lambda_1 \mid t)$ невозможной. Здесь предполагается, что значение T известно точно.

В [1] сформулирован алгоритм расчета апостериорной вероятности $w(\lambda_1 \mid t)$ для случая отсутствия мертвого времени ($T = 0$). При этом поведение вероятности

$w(\lambda_1 \mid t)$ на полуинтервале $[t_i, t_{i+1})$ между соседними наблюдаемыми событиями обобщенного полусинхронного потока определяются выражением (1)

$$w(\lambda_1 \mid t) = \frac{w[1 - w(\lambda_1 \mid t_i + 0)] - [w - w(\lambda_1 \mid t_i + 0)]e^{-b(t-t_i)}}{1 - w(\lambda_1 \mid t_i + 0) - [w - w(\lambda_1 \mid t_i + 0)]e^{-b(t-t_i)}} \tag{1}$$

где $t_i \le t < t_{i+1}$, $i = 0, 1, \ldots$; $b = \lambda_1 - \lambda_2 - \alpha$; $w = \alpha(1-\delta)/(\lambda_1 - \lambda_2 - \alpha\delta) \ne 0$, $0 \le \delta \le 1$.

В момент времени t_i (в момент наступления события обобщенного полусинхронного потока) апостериорная вероятность (1) претерпевает разрыв 1-го рода ($i = 1, 2, \ldots$), поэтому в момент времени t_i имеет место формула пересчета (2):

$$w(\lambda_1 \mid t_i + 0) = \frac{\alpha\delta + [(1-p)\lambda_1 - \alpha\delta]w(\lambda_1 \mid t_i - 0)}{\lambda_2 + \alpha\delta + (\lambda_1 - \lambda_2 - \alpha\delta)w(\lambda_1 \mid t_i - 0)}, i = 1, 2, \ldots. \tag{2}$$

где $w(\lambda_1 \mid t_i - 0)$ вычисляется по формуле (1) в момент времени $t = t_i$, когда t изменяется в полуинтервале $[t_{i-1}, t_i)$, соседнем с полуинтервалом $[t_i, t_{i+1})$. В качестве начального значения $w(\lambda_1 \mid t_0 + 0) = w(\lambda_1 \mid t_0 = 0)$ в (1) выбирается априорная финальная вероятность первого состояния процесса $\lambda(t)$: $\pi_1 = \alpha/(\alpha + p\lambda_1)$, которая находится из уравнений $p\lambda_1\pi_1 - \alpha\pi_2 = 0$, $\pi_1 + \pi_2 = 1$.

Таким образом, вычисление апостериорной вероятности $w(\lambda_1 \mid t)$ по формуле (1) в условиях, когда длительность мертвого времени $T \ne 0$, справедливо на интервале $(t_i + T, t_{i+1})$, значение длительности которого есть η_i. При этом начальное условие для $w(\lambda_1 \mid t)$ привязывается к моменту времени $t_i + T$, т.е. в формуле (1), во-первых, нужно $w(\lambda_1 \mid t_i{+}0)$ заменить на $w(\lambda_1 \mid t_i{+}T)$, во-вторых, $t_i{+}T \le t < t_{i+1}$, $i = 1, 2, \ldots$. Формула пересчета (2) остается при этом без изменения, так как она предназначена для вычисления апостериорной вероятности в момент времени t_i наступления наблюдаемого события, которое порождает мертвое время.

Можно показать, что на интервале $(t_i, t_i + T)$ апостериорные вероятности $w(\lambda_j \mid t)$ удовлетворяют системе дифференциальных уравнений:

$$\frac{dw(\lambda_1 \mid t)}{dt} = -p\lambda_1 w(\lambda_1 \mid t) + \alpha w(\lambda_2 \mid t), \frac{dw(\lambda_2 \mid t)}{dt} = p\lambda_1 w(\lambda_1 \mid t) - \alpha w(\lambda_2 \mid t), \tag{3}$$

с граничными условиями: $w(\lambda_1 \mid t = t_i) = w(\lambda_1 \mid t_i + 0)$, $w(\lambda_2 \mid t = t_i) = w(\lambda_2 \mid t_i + 0)$, $i = 1, 2, \ldots$. Последнее вытекает из того, что на интервале $(t_{i-1} + T, t_i)$, $i = 2, 3, \ldots$, смежным интервалу $(t_i, t_i + T)$ апостериорная вероятность рассчитывается по формуле (1), где вместо $w(\lambda_1 \mid t_i{+}0)$ стоит $w(\lambda_1 \mid t_i{+}T)$; в точке $t = t_i$ происходит пересчет апостериорной вероятности по формуле (2), так что её значение в этой точке есть $w(\lambda_1 \mid t_i + 0)$. Для граничного интервала (t_0, t_1) расчет апостериорной вероятности $w(\lambda_1 \mid t)$ производится по формуле (1) с её последующим пересчетом по формуле (2) в точке $t = t_1$. Решая систему (3), находим

$$w(\lambda_1 \mid t) = \pi_1 + (w(\lambda_1 \mid t_i + 0) - \pi_1)e^{-(\alpha + p\lambda_1)(t - t_i)}, \tag{4}$$

где $t_i \le t \le t_i + T$, $i = 1, 2, \ldots$; вероятность π_1 определена в (1). Тогда из (4) следует, что

$$w(\lambda_1 \mid t_i + T) = \pi_1 + (w(\lambda_1 \mid t_i + 0) - \pi_1)e^{-(\alpha + p\lambda_1)T}, i = 1, 2, \ldots. \quad (5)$$

Полученные формулы позволяют сформулировать алгоритм расчета апостериорной вероятности $w(\lambda_1 \mid t)$ и алгоритм принятия решения о состоянии процесса $\lambda(t)$ в любой момент времени t: 1) в момент времени $t_0 = 0$ задается $w(\lambda_1 \mid t_0 + 0) = w(\lambda_1 \mid t_0 = 0) = \pi_1$; 2) по формуле (1) для $i = 0$ рассчитывается вероятность $w(\lambda_1 \mid t)$ в любой момент времени t $(0 \le t < t_1)$, где t_1 - момент наблюдения первого события наблюдаемого потока; 3) по формуле (1) для $i = 0$ рассчитывается вероятность $w(\lambda_1 \mid t)$ в момент времени $t_1 : w(\lambda_1 \mid t_1) = w(\lambda_1 \mid t_1 - 0)$; 4) i увеличивается на единицу и по формуле (2) для $i = 1$ производится пересчет апостериорной вероятности $w(\lambda_1 \mid t)$ в момент времени $t = t_1$, при этом $w(\lambda_1 \mid t_1 + 0)$ является начальным значением для $w(\lambda_1 \mid t)$ в формуле (4); 5) по формуле (4) для $i = 1$ рассчитывается вероятность $w(\lambda_1 \mid t)$ в любой момент времени t $(t_1 < t < t_1 + T)$; 6) по формуле (5) для $i = 1$ рассчитывается вероятность $w(\lambda_1 \mid t)$ в момент времени $t = t_1 + T$, т.е. $w(\lambda_1 \mid t_1 + T)$; при этом $w(\lambda_1 \mid t_1 + T)$ является начальным условием для $w(\lambda_1 \mid t)$ на следующем шаге алгоритма;

7) для $i = 1$ по формуле

$$w(\lambda_1 \mid t) = \frac{w[1 - w(\lambda_1 \mid t_i + T)] - [w - w(\lambda_1 \mid t_i + T)]e^{-b(t - t_i - T)}}{1 - w(\lambda_1 \mid t_i + T) - [w - w(\lambda_1 \mid t_i + T)]e^{-b(t - t_i - T)}} \quad (6)$$

$(t_i + T \le t < t_{i+1}$, $i = 1, 2, \ldots$; w, b определены в (1)) рассчитывается вероятность $w(\lambda_1 \mid t)$ в любой момент времени t $(t_1 + T < t < t_2)$, где t_2 - момент наблюдения второго события наблюдаемого потока; 8) по формуле (6) для $i = 1$ рассчитывается вероятность $w(\lambda_1 \mid t)$ в момент времени $t = t_2 : w(\lambda_1 \mid t_2) = w(\lambda_1 \mid t_2 - 0)$; 9) алгоритм переходит на шаг 4, после чего шаги 4-8 повторяются для $i = 2$ и т.д.

Параллельно по ходу вычисления апостериорной вероятности $w(\lambda_1 \mid t)$ в любой момент времени t выносится решение о состоянии процесса $\lambda(t)$: если $w(\lambda_1 \mid t) \ge w(\lambda_2 \mid t)$ $(w(\lambda_1 \mid t) \ge 1/2)$, то оценка $\widehat{\lambda}(t) = \lambda_1$, в противном случае $\widehat{\lambda}(t) = \lambda_2$.

4. ЗАКЛЮЧЕНИЕ

Полученные путем имитационного моделирования экспериментальные результаты показывают возможность оценивания состояний обобщенного полусинхронного потока событий в условиях непродлевающегося мертвого времени по результатам текущих наблюдений (в течение некоторого временного интервала) за потоком. Это, в свою очередь, позволяет изменять режимы функционирования системы массового обслуживания в зависимости от того или иного состояния обобщенного полусинхронного потока событий (адаптироваться к изменяющейся интенсивности входящего потока событий). Выражения апостериорных вероятностей для оценки состояний потока получены в явном виде, что позволяет производить вычисления без привлечения численных методов.

ЛИТЕРАТУРА

1. *Горцев А.М., Калягин А.А., Нежельская Л.А.* Оптимальная оценка состояний обобщенного полусинхронного потока событий // Вестник ТГУ. Управление, вычислительная техника и информатика. 2010. №2(11). С. 66-81.
2. *Дудин А.Н., Клименок В.И.* Системы массового обслуживания с коррелированными потоками. Мн.: Издательство БГУ, 2000. 175 С.
3. *Горцев А.М., Леонова М.А.* Оптимальная оценка состояний обобщенного асинхронного дважды стохастического потока // Вестник ТГУ. Управление, вычислительная техника и информатика. 2010. №1(10). С. 33-47.
4. *Горцев А.М., Зуевич В.Л.* Оптимальная оценка состояний асинхронного дважды стохастического потока событий с произвольным числом состояний // Вестник ТГУ. Управление, вычислительная техника и информатика. 2010. №2(11). С. 44-65.
5. *Апанасович В.В., Коляда А.А., Чернявский А.Ф.* Статистический анализ случайных потоков в физическом эксперименте. Мн.: Издательство "Университетское 1988. 254 С.
6. *Горцев А.М., Климов И.С.* Оценка интенсивности пуассоновского потока событий в условиях частичной его ненаблюдаемости // Радиотехника. 1991. №12. С. 3-7.
7. *Горцев А.М., Климов И.С.* Оценивание параметров знакопеременного пуассоновского потока событий // Радиотехника. 1994. №8. С. 3-9.
8. *Хазен Э.М.* Методы оптимальных статистических решений и задачи оптимального управления. М.: Сов. радио, 1968. 256 С.

ДВУХФАЗНАЯ СИСТЕМА ОБСЛУЖИВАНИЯ С ОЖИДАНИЕМ И ДОПОЛНИТЕЛЬНЫМ ПОТОКОМ НА ВТОРУЮ ФАЗУ

Ч.С. Ким[1], В. Клименок[2,*], О. Тарамин[2]

[1] *Университет Сангжи,*

[2] *Белорусский государственный университет*

[1] *Вонжу, Корея*

[2] *Минск, Беларусь*

[*] vklimenok@yandex.ru

В статье исследуется двухфазная система с ожиданием на первой фазе, марковским входным потоком неприоритетных запросов и дополнительным стационарным пуассоновским потоком приоритетных запросов, поступающих непосредственно на вторую фазу. Предполагается, что неприоритетный запрос после прохождения обработки на первой фазе может занимать несколько приборов второй фазы. В системе предусмотрены как потери неприоритетного запроса после обслуживания на первой фазе, так и блокировки обслуживающего прибора первой фазы этим запросом в случае недоступности необходимого числа приборов на второй фазе.

Ключевые слова: двухфазная система обслуживания, марковский входной поток, дополнительный стационарный пуассоновский поток.

1. ВВЕДЕНИЕ

Двухфазные системы массового обслуживания являются математическими моделями двух последовательных узлов сети, которые позволяют учесть взаимное влияние процессов обслуживания в этих узлах. Данные системы являются важным частным случаем сетей массового обслуживания и служат адекватными математическими моделями фрагментов телекоммуникационных сетей. Они широко используются при проектировании, анализе функционирования и оптимизации информационно-телекоммуникационных сетей различного назначения, а также для проверки алгоритмов декомпозиции сетей общего вида.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается двухфазная система, первую фазу которой образует однолинейная система массового обслуживания с ожиданием. В систему поступает марковский входной поток (MAP – Markovian Arrival Process) запросов, заданный

неприводимой цепью Маркова ν_t, $t \geq 0$, с непрерывным временем и конечным пространством состояний $\{0, ..., W\}$. Интенсивности переходов процесса ν_t, $t \geq 0$, сопровождающихся поступлением запроса (не сопровождающихся поступлением запроса), задаются элементами матрицы D_1 (недиагональными элементами матрицы D_0). Запросы MAP-потока будем называть *неприоритетными.* Интенсивность поступления неприоритетных запросов задается формулой $\lambda_1 = \boldsymbol{\theta} D_1 \mathbf{e}$, где $\boldsymbol{\theta}$ является единственным решением системы линейных алгебраических уравнений $\boldsymbol{\theta}(D_0 + D_1) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$. Здесь и далее $\mathbf{e}$ – вектор-столбец, состоящий из единиц, $\mathbf{0}$ – вектор-строка, состоящая из нулей.

Времена обслуживания запросов на первой фазе являются независимыми случайными величинами с функцией распределения $B(t)$ и конечным первым моментом $b_1 = \int\limits_0^\infty t dB(t)$.

После обслуживания на первой фазе неприоритетный запрос поступает на вторую фазу рассматриваемой системы, которую образуют N независимых идентичных приборов.

Для обслуживания на второй фазе неприоритетному запросу требуется случайное число приборов ϕ. Здесь ϕ – целочисленная случайная величина с распределением $q_m = P\{\phi = m\}$, $q_m \geq 0$, $m = \overline{0, N}$, $\sum\limits_{m=0}^{N} q_m = 1$. Если в момент окончания обслуживания запроса прибором первой фазы необходимое число свободных приборов на второй фазе отсутствует, то с вероятностью p, $0 \leq p \leq 1$, запрос уходит из системы недообслуженным (теряется), а с дополнительной вероятностью ждет, пока освободится нужное число приборов. Период ожидания сопровождается блокировкой прибора первой фазы.

Кроме неприоритетных запросов, дополнительный стационарный пуассоновский поток с интенсивностью λ_2 *приоритетных* запросов поступает непосредственно на вторую фазу. Если приоритетный запрос застает один из приборов второй фазы свободным, то этот запрос начинает обслуживание, в противном случае он покидает систему (теряется). Для обслуживания на второй фазе приоритетному запросу требуется один прибор.

После занятия запросом приборов второй фазы каждый из приборов независимо от других обслуживает этот запрос в течение экспоненциально распределенного времени с параметром μ.

3. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВЛОЖЕННОЙ ЦЕПИ МАРКОВА

Пусть t_n – n-й момент окончания обслуживания на первой фазе, $n \geq 1$. Рассмотрим процесс $\xi_n = \{i_n, r_n, \nu_n\}$, $n \geq 1$, где i_n – число запросов на первой фазе (не включая запрос, вызвавший блокировку) в момент времени $t_n + 0$; r_n – число занятых приборов на второй фазе в момент времени $t_n - 0$; ν_n – состояние MAP-потока в момент времени t_n, $i_n \geq 0, r_n = \overline{0, N}, \nu_n = \overline{0, W}$.

Процесс ξ_n, $n \geq 1$, является неприводимой непериодической цепью Маркова. Перенумеруем все состояния данной цепью в лексикографическом порядке и сформируем матрицы вероятностей переходов $P_{i,j}$, $i, j \geq 0$, которые задают вероятности переходов цепи из состояний, соответствующих значению i счетной компоненты, в состояния, соответствующие значению j этой компоненты.

Введем следующие обозначения:

- O (I) – нулевая (единичная) матрица, $R = \mathrm{diag}\{r, r = \overline{0, N}\}$,$\bar{W} = W + 1$;
- $\otimes$ $(\oplus)$ – символ кронекерова произведения (кронекеровой суммы) матриц;
- матрицы $P(n,t), n \geq 0,$ – коэффициенты разложения $e^{(D_0+D_1 z)t} = \sum\limits_{n=0}^{\infty} P(n,t) z^n$;
- $F(t) = (F_{r,r'}(t))_{r,r'=\overline{0,N}}$, где $F_{r,r'}(t) = 0$, если $r \leq r'$, и, если $r > r'$, $F_{r,r'}(t)$ – функция распределения времени блокировки при условии, что в начале и в конце блокировки заняты r и r' приборов второй фазы соответственно;
- $Q_2 = \mathrm{diag}\{\sum\limits_{m=N-r+1}^{N} q_m, r = \overline{0, N}\}$,

$$Q_1 = \begin{pmatrix} q_0 & q_1 & \dots & q_N \\ 0 & q_0 & \dots & q_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_0 \end{pmatrix}, Q_3 = \begin{pmatrix} 0 & \dots & 0 & q_N \\ 0 & \dots & 0 & q_{N-1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & q_0 \end{pmatrix};$$

- $\tilde{Q}_m = Q_m \otimes I_{\bar{W}}$, $m = \overline{1, 3}$;
- $\bar{Q} = \tilde{Q}_1 + p\tilde{Q}_2$, $\hat{Q} = \tilde{Q}_1 + p\tilde{Q}_2 + (1-p)\int\limits_0^{\infty} (dF(t) \otimes e^{D_0 t})\tilde{Q}_3$.

Лемма 1. *Матрица вероятностей переходов цепи Маркова $\xi_n, n \geq 1$, имеет блочную структуру*

$$P = (P_{i,j})_{i,j \geq 0} = \begin{pmatrix} V_0 & V_1 & V_2 & V_3 & \cdots \\ Y_0 & Y_1 & Y_2 & Y_3 & \cdots \\ 0 & Y_0 & Y_1 & Y_2 & \cdots \\ 0 & 0 & Y_0 & Y_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

где

$$V_i = -\hat{Q}(\Gamma \oplus D_0)^{-1}(I_{N+1} \otimes D_1)\Omega_i + (1-p)\sum_{k=1}^{i+1} \mathcal{F}_k \tilde{Q}_3 \Omega_{i-k+1},$$

$$Y_i = \bar{Q}\Omega_i + (1-p)\sum_{k=0}^{i} \mathcal{F}_k \tilde{Q}_3 \Omega_{i-k},$$

$$\Omega_n = \int\limits_0^{\infty} e^{\Gamma t} \otimes P(n,t) dB(t), \qquad \mathcal{F}_n = \int\limits_0^{\infty} dF(t) \otimes P(n,t),\ n \geq 0,$$

$$\Gamma = \begin{pmatrix} -\lambda_2 & \lambda_2 & 0 & 0 & \cdots & 0 & 0 \\ \mu & -\lambda_2-\mu & \lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & 2\mu & -\lambda_2-2\mu & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & N\mu & -N\mu \end{pmatrix}.$$

Следствие 1. *Процесс $\xi_n, n \geq 1$, является многомерной квазитеплицевой цепью Маркова.*

Следствие 2. *Матричные производящие функции $V(z)$ и $Y(z)$ имеют вид*

$$V(z) = [-\hat{Q}(\Gamma \oplus D_0)^{-1}(I_{N+1} \otimes D_1) + \frac{1}{z}(1-p)(\mathcal{F}(z) - \mathcal{F}_0)\tilde{Q}_3]\Omega(z),$$

$$Y(z) = [\bar{Q} + (1-p)\mathcal{F}(z)\tilde{Q}_3]\Omega(z),$$

где $\Omega(z) = \int\limits_0^\infty e^{\Gamma t} \otimes e^{(D_0+D_1 z)t} dB(t)$, $\mathcal{F}(z) = \int\limits_0^\infty dF(t) \otimes e^{(D_0+D_1 z)t}$, $|z| \leq 1$.

Теорема 1. *Необходимым и достаточным условием существования стационарного распределения цепи Маркова ξ_n, $n \geq 1$, является выполнение неравенства*

$$\rho = \lambda[b_1 + (1-p)\sum_{r=1}^{N} y_r \sum_{m=N-r+1}^{N} q_m \sum_{l=N-m+1}^{r} (l\mu)^{-1}] < 1.$$

Здесь $(y_1, ..., y_N)$ – часть вектора $\boldsymbol{y} = (y_0, y_1, ..., y_N)$, который является единственным решением системы

$$\boldsymbol{y}\mathcal{Q} \int\limits_0^\infty e^{\Gamma t} dB(t) = \boldsymbol{y}, \quad \boldsymbol{y}\mathbf{e} = 1,$$

где $\mathcal{Q} = Q_1 + pQ_2 + (1-p)EQ_3$, E – матрица, у которой все элементы нулевые, кроме поддиагональных, равных единице.

Пусть $\pi(i, r, \nu)$, $i \geq 0, r = \overline{0, N}$, $\nu = \overline{0, W}$, – стационарное распределение цепи Маркова ξ_n, $n \geq 1$. Сформируем вектор-строки $\boldsymbol{\pi}_i$ стационарных вероятностей $\pi(i, r, \nu)$, сгруппированных в лексикографическом порядке компонент (r, ν), $i \geq 0$. Пусть $\mathbf{\Pi}(z) = \sum\limits_{i=0}^{\infty} \boldsymbol{\pi}_i z^i$, $|z| \leq 1$, – производящая функция векторов $\boldsymbol{\pi}_i$, $i \geq 0$.

Для вычисления векторов $\boldsymbol{\pi}_i$, $i \geq 0$, используется численно устойчивый алгоритм, разработанный в [1] для нахождения стационарного распределения многомерных квазитеплицевых цепей Маркова.

4. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ В ПРОИЗВОЛЬНЫЙ МОМЕНТ ВРЕМЕНИ

Рассмотрим процесс $\zeta_t = \{i_t, k_t, r_t, \nu_t\}$, $t \geq 0$, состояний системы в произвольный момент времени, где i_t – число запросов на первой фазе (включая запрос,

вызвавший блокировку); k_t – величина, принимающая значения 0, 1, 2 в зависимости от того, свободен, обслуживает запрос или блокирован прибор первой фазы; r_t – число занятых приборов на второй фазе; ν_t – состояние MAP-потока в момент времени t, $t \geq 0$.

Пусть $p(i,k,r,\nu) = \lim\limits_{t\to\infty} P\{i_t = i, k_t = k, r_t = r, \nu_t = \nu\}$, $i \geq 0$, $k = 0,1,2, r = \overline{0,N}$, $\nu = \overline{0,W}$, – стационарное распределение процесса ζ_t, $t \geq 0$, и $\boldsymbol{p}_i(k)$ – векторы этих вероятностей, упорядоченных в лексикографическом порядке.

Теорема 2. *Ненулевые векторы стационарных вероятностей* $\boldsymbol{p}_i(k)$, $i \geq 0$, $k = 0,1,2$, *выражаются через векторы стационарного распределения* $\boldsymbol{\pi}_i$, $i \geq 0$, *вложенной цепи Маркова* ξ_n, $n \geq 1$, *следующим образом:*

$$\boldsymbol{p}_0(0) = -\tau^{-1}\boldsymbol{\pi}_0\hat{Q}(\Gamma \oplus D_0)^{-1},$$

$$\boldsymbol{p}_i(1) = \tau^{-1}\{\boldsymbol{\pi}_0[-\hat{Q}(\Gamma \oplus D_0)^{-1}(I_{N+1} \otimes D_1)\tilde{\Omega}_i + (1-p)\sum_{k=1}^{i}\mathcal{F}_k\tilde{Q}_3]\tilde{\Omega}_{i-k} +$$

$$+\sum_{l=1}^{i}\boldsymbol{\pi}_l[\bar{Q}\tilde{\Omega}_{i-l} + (1-p)\sum_{k=0}^{i-l}\mathcal{F}_k\tilde{Q}_3\tilde{\Omega}_{i-k-l}]\},$$

$$\boldsymbol{p}_i(2) = \tau^{-1}(1-p)\sum_{l=0}^{i-1}\boldsymbol{\pi}_l\sum_{k=0}^{i-l-1}(\mathcal{F}_k + \delta_{0,k}I)\tilde{Q}_2\int\limits_0^\infty e^{-\mu Rt} \otimes P(i-k-l-1,t)dt,\ i \geq 1,$$

где τ *– средняя величина интервала между двумя соседними моментами окончания обслуживания на первой фазе,*

$$\tau = b_1 - \boldsymbol{\pi}_0\hat{Q}(I_{N+1} \otimes D_0)^{-1}\mathbf{e} + (1-p)\mathbf{\Pi}(1)(I_{N+1} \otimes \mathbf{e})\int\limits_0^\infty tdF(t)Q_3\mathbf{e},$$

$\tilde{\Omega}_n = \int\limits_0^\infty e^{\Gamma t} \otimes P(n,t)(1-B(t))dt$, $n \geq 0$, $\delta_{i,j}$ *– символ Кронекера.*

Следствие 3. *Векторы стационарных вероятностей* $\boldsymbol{p}_i$, $i \geq 0$, *процесса* $\{i_t, r_t, \nu_t\}$, $t \geq 0$, *вычисляются следующим образом:* $\boldsymbol{p}_i = \sum\limits_{k=0}^{2}\boldsymbol{p}_i(k)$, $i \geq 0$.

Следствие 4. *Производящая функция* $\boldsymbol{P}(z) = \sum\limits_{i=0}^{\infty}\boldsymbol{p}_i z^i$, $|z| \leq 1$, *выражается через производящую функцию* $\mathbf{\Pi}(z)$ *следующим образом:*

$$\boldsymbol{P}(z)(\Gamma \oplus (D_0 + zD_1)) = \tau^{-1}\mathbf{\Pi}(z)\{z[I - (1-p)\mathcal{F}(z)\tilde{Q}_2 \times$$

$$\times(\mu R \oplus (-D_0 - zD_1))^{-1}(\Gamma \oplus (D_0 + zD_1))] - [\bar{Q} + (1-p)\mathcal{F}(z)\tilde{Q}_3]\}.$$

5. ХАРАКТЕРИСТИКИ ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМЫ

Среднее число запросов на первой фазе в момент окончания обслуживания на этой фазе и в произвольный момент времени:

$$L = \mathbf{\Pi}'(1)\mathbf{e}, \quad \tilde{L} = \boldsymbol{P}'(1)\mathbf{e}.$$

Среднее число занятых приборов на второй фазе в моменты окончания обслуживания на первой фазе и в произвольный момент времени:

$$N_{busy} = \mathbf{\Pi}(1)(I_{N+1} \otimes \mathbf{e}_{\bar{W}})R\mathbf{e}, \qquad \tilde{N}_{busy} = \boldsymbol{P}(1)(I_{N+1} \otimes \mathbf{e}_{\bar{W}})R\mathbf{e}.$$

Вероятность того, что непроизвольный приоритетный запрос покинет систему (вызовет блокировку) после обслуживания на первой фазе:

$$P_{loss}^{(1)} = p\mathbf{\Pi}(1)\tilde{Q}_2\mathbf{e}, \quad P_{block} = 1 - P_{loss}^{(1)}.$$

Вероятность того, что произвольный приоритетный запрос будет потерян:

$$P_{loss}^{(2)} = \mathbf{\Pi}(1)(\hat{\mathbf{e}}_{N+1} \otimes \mathbf{e}_{\bar{W}}),$$

где $\hat{\mathbf{e}}_{N+1}$ – вектор-столбец размерности $N+1$, последняя компонента которого равна 1, а остальные – нулевые.

Вероятности P_{idle}, P_{serve}, P_{block} того, что прибор первой фазы свободен, занят обслуживанием или заблокирован:

$$P_{idle} = -\tau^{-1}\boldsymbol{\pi}_0\hat{Q}(I_{N+1} \otimes D_0)^{-1}\mathbf{e}, \; P_{serve} = \tau^{-1}b_1, \; P_{block} = 1 - P_{idle} - P_{serve}.$$

6. ЗАКЛЮЧЕНИЕ

В данной статье исследована двухфазная система обслуживания с ожиданием и дополнительным потоком на вторую фазу. Функционирование системы описано в терминах многомерной квазитеплицевой цепи Маркова и полурегенерирующего процесса, получено необходимое и достаточное условие существования стационарного режима, найдены стационарные распределения вероятностей системы во вложенные моменты и в произвольный момент времени, получены формулы для расчета основных характеристик производительности.

ЛИТЕРАТУРА

1. *Dudin A.N., Klimenok V.I.* Multi-dimensional quasi-Toeplitz Markov chains // Mathematics and Stochastic Analysis. 1999. V. 12. P. 393–415.
2. *Cinlar E.* Introduction to stochastic process // N.J.: Prentice-Hall. 1975.

APPLICATION OF MULTILEVEL HM-NETWORK AT DESIGNING OF WAREHOUSES SQUARES IN LOGISTICS TRANSPORT SYSTEMS

O. Kiturko

Grodno State University
Grodno, Belarus
sytaya_om@mail.ru

The article deals with the technique allowing to estimate and forecast expected incomes, logistics transport systems subjects warehouse squares. The technique is based on application of HM (Howard–Matalytski) – queueing networks.

Keywords: HM – queueing networks, logistics transport systems, warehouse squares.

1. INTRODUCTION

The logistics structure is a system which consists of various functional areas, such as: stores, information, stacking and warehouse handling, transporting of products and other areas [1]. Thus the main task is cost minimization and profit maximization for producers', customers', warehouses' etc.

Working out LS models is an important problem. The number and the arrangement of producing units (the enterprises, firms, etc.), an amount and the arrangement of warehouses, transport models, connection and information systems are to be taken into consideration.

2. APPLICATION OF NETWORKS AT WAREHOUSE DESIGNING

We'd like to mention that the HM-network can be used not only for LTS subjects expected incomes forecasting, but also for designing of warehouse squares, determination of transport warehouse workers amount in brigades that are engaged in cargo loading and unloading. We will illustrate it in the following example. Let's assume that LTS subjects are n warehouses S_1, S_2, ..., S_n between which cargo transportation is carried out. For cars loading and unloading in a warehouse S_i m_i brigades are created, $i = \overline{1, n}$. For simplicity we suggest that the same brigade is engaged in car unloading and in loading it afterwards with new production for further transportation so that the car's stay idle time was minimum. Transporting of the goods from one subject to another brings the latter some casual income and consequently the income of the first subject is reduced to this random variable (RV), however, or it can be vice versa, it doesn't matter for the model. Let's consider the dynamics of the network system S_i

incomes modification (warehouse S_i of LTS). Let's designate by $V_i(t)$ system S_i income at the moment t, and by $v_{i0} = V_i(0)$ its income at the initial moment. Then it is possible to present its income at the moment $t + \Delta t$ as the following:

$$V_i(t + \Delta t) = V_i(t) + \Delta V_i(t, \Delta t), \tag{1}$$

where $\Delta V_i(t, \Delta t)$ – the modification of income QS S_i inn time interval $[t, t + \Delta t)$. To determinate this magnitude let's write out probable events which can happen in time Δt, and incomes modification of S_i systems connected with these events.

1) With the probability $\lambda(t)p_{0i}\Delta t + o(\Delta t)$ the request to system S_i will be received (to a warehouse S_i car) which will bring some income, occupying space r_{0i}, where r_{0i} – RV with expectation $M\{r_{0i}\} = a_{0i}$, $i = \overline{1, n}$.

2) With the probability $\mu_i(k_i(t))u(k_i(t))p_{i0}\Delta t + o(\Delta t)$ the request of S_i will go to the outer world, thus income of QS S_i will be reduced to the magnitude, occupying the of space R_{i0}, where R_{i0} – RV from expectation $M\{R_{i0}\} = b_{i0}$, $i = \overline{1, n}$.

3) With the probability $\mu_j(k_j(t))u(k_j(t))p_{ji}\Delta t + o(\Delta t)$ the request S_j system will pass to the S_i system, thus S_i system income will increase by magnitude, occupying the space of r_{ji}, and the S_j income (i.e. the occupied area of the warehouse S_j) will decrease by this magnitude, where r_{ji} – RV from expectation $M\{r_{ji}\} = a_{ji}$, $i, j = \overline{1, n}$, $i \neq j$.

4) With the probability $\mu_i(k_i(t))u(k_i(t))p_{ij}\Delta t + o(\Delta t)$ a request from system S_i transits to system S_j, thus the income of S_i will decrease by magnitude, occupying space R_{ij}, and the income of S_j (i.e. the occupied square of the warehouse S_j) will increase by this magnitude, where R_{ij} – RV from expectation $M\{R_{ij}\} = b_{ij}$, $i, j = \overline{1, n}$, $i \neq j$.

5) With the probability $1 - (\lambda(t)p_{0i} + \mu_i(k_i(t))u(k_i(t)) + \sum\limits_{\substack{j=1,\\ j\neq i}}^{n} \mu_j(k_j(t))u(k_j(t))p_{ji}) \times$ $\times \Delta t + o(t)$ at time interval $[t, t + \Delta t)$ the S_i system state modification will not happen (i.e. the occupied square of the warehouse S_i will not vary), $i = \overline{1, n}$.

Besides, for each small time interval Δt system S_i (subject S_i) increases the income by magnitude $r_i\Delta t$ by the means of percent on the money which is in its bank, where r_i – RV from expectation $M\{r_i\} = c_i$, $i = \overline{1, n}$. We will consider also that RV r_{ji}, R_{ij}, r_{0i}, R_{i0} are independent in relation to RV r_i, $i, j = \overline{1, n}$.

It is obvious that $r_{ji} = R_{ji}$ with probability 1, i.e.

$$a_{ji} = b_{ji}, \quad i, j = \overline{1, n}. \tag{2}$$

Then from the mentioned above we can conclude the following, at the fixed realization of process $k(t)$, it is possible to note:

$$M\{\Delta V_i(t, \Delta t)\} = \sum_k P(k(t) = k)\, M\{\Delta V_i(t, \Delta t)/k(t)\} =$$

$$= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \ldots \sum_{k_n=0}^{\infty} P(k(t) = (k_1(t), k_2(t), \ldots, k_n(t))) \times$$

$$\times M\{\Delta V_i(t,\,\Delta t)/k(t) = (k_1(t),\,k_2(t),\,...,\,k_n(t))\} =$$

$$= [\lambda(t)p_{0i}a_{0i} + c_i - (p_{i0}b_{i0} + \sum_{\substack{j=1\\ j\neq i}}^{n} p_{ij}b_{ij})\sum_k P\,(k(t)=k)\mu_i(k_i(t))u(k_i(t)) +$$

$$+ \sum_{\substack{j=1\\ j\neq i}}^{n} p_{ji}a_{ji}\sum_k P\,(k(t)=k)\mu_j(k_j(t))u(k_j(t))]\Delta t + o(\Delta t).$$

Averaging on $k(t)$ taking into account a normalization state $\sum_k P\,(k(t)=k) = 1$, for modification of subject S_i expected income we receive:

$$M\{\Delta V_i(t,\,\Delta t)\} = \sum_k P\,(k(t)=k)\,M\{\Delta V_i(t,\,\Delta t)/k(t)\} =$$

$$= \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}...\sum_{k_n=0}^{\infty} P\,(k(t) = (k_1(t),\,k_2(t),\,...,\,k_n(t)))\times$$

$$\times M\{\Delta V_i(t,\,\Delta t)/k(t) = (k_1(t),\,k_2(t),\,...,\,k_n(t))\} =$$

$$= [\lambda(t)p_{0i}a_{0i} + c_i - (p_{i0}b_{i0} + \sum_{\substack{j=1\\ j\neq i}}^{n} p_{ij}b_{ij})\sum_k P\,(k(t)=k)\mu_i(k_i(t))u(k_i(t)) +$$

$$+ \sum_{\substack{j=1\\ j\neq i}}^{n} p_{ji}a_{ji}\sum_k P\,(k(t)=k)\mu_j(k_j(t))u(k_j(t))]\Delta t + o(\Delta t).$$

Let's consider that time intervals of servicing a request in the system S_i (intervals of one car "unloadings – loadings" at the warehouse S_i) are distributed upon the demonstrative law with parameter μ_i, $i = \overline{1,\,n}$, and $\lambda(t) = \lambda$. In this case

$$\mu_i(k_i(t)) = \begin{cases} \mu_i k_i(t), & k_i(t) \le m_i, \\ \mu_i m_i, & k_i(t) > m_i, \end{cases} \qquad \mu_i(k_i(t))u(k_i(t)) = \mu_i \min(k_i(t),\,m_i), \quad i = \overline{1,\,n}.$$

Let's assume also that the expression averaging $\mu_i(k_i(t))u(k_i(t))$ gives $\mu_i \min(N_i(t),\,m_i)$, i.e.

$$M\min(k_i(t),\,m_i) = \min(N_i(t),\,m_i), \tag{3}$$

where $N_i(t)$ is the average number of requests (expecting and served) in S_i at the moment of time t, $i = \overline{1,\,n}$. Taking into account this supposition we receive the following approximate relation:

$$M\{\Delta V_i(t,\,\Delta t)\} = [\lambda(t)p_{0i}a_{0i} + c_i - \mu_i \min(N_i(t),\, m_i)(p_{i0}b_{i0} + \sum_{\substack{j=1\\ j\neq i}}^{n} p_{ij}b_{ij}) +$$

$$+ \sum_{\substack{j=1\\ j\neq i}}^{n} \mu_j \min(N_j(t),\, m_j)p_{ji}a_{ji}]\Delta t + o(\Delta t). \quad (4)$$

As an elementary stream of requests arrives in a network with intensity λ, i.e. the probability of a requests inflow in S_i for time Δt looks like $P_a(\Delta t) = \dfrac{(\lambda p_{0i}\Delta t)^a}{a!}e^{-\lambda p_{0i}\Delta t}$, $a = 0,\, 1,\, 2,\, ...$, the average number of the requests which have arrived from the outside to S_i for time Δt is equal to $\lambda p_{0i}\Delta t$. We will find the average number of the occupied service lines in S_i at the moment t, $i = \overline{1,\, n}$, by $\rho_i(t)$. Then $\mu_i\rho_i(t)\Delta t$ is the average number of the requests which have abandoned S_i for time Δt, and $\sum\limits_{\substack{j=1\\ j\neq i}}^{n} \mu_j\rho_j(t)p_{ji}\Delta t$ is the average number of the requests which have arrived to S_i from other subjects for time Δt. Therefore

$$N_i(t+\Delta t) - N_i(t) = \lambda p_{0i}\Delta t + \sum_{\substack{j=1\\ j\neq i}}^{n} \mu_j\rho_j(t)p_{ji}\Delta t - \mu_i\rho_i(t)\Delta t, \quad i = \overline{1,\, n},$$

Consequently, at the $\Delta t \to 0$ we receive the UDE system for $N_i(t)$:

$$\frac{dN_i(t)}{dt} = \sum_{\substack{j=1\\ j\neq i}}^{n} \mu_j\rho_j(t)p_{ji} - \mu_i\rho_i(t) + \lambda p_{0i}, \quad i = \overline{1,\, n}. \quad (5)$$

The precise magnitude $\rho_i(t)$ is impossible to discover and consequently it is approximated by the expression

$$\rho_i(t) = \begin{cases} N_i(t),\ N_i(t) \le m_i, \\ m_i,\ N_i(t) > m_i, \end{cases} = \min(N_i(t),\ m_i)\ .$$

Then the set of equations (5) will become

$$\frac{dN_i(t)}{dt} = \sum_{\substack{j=1\\ j\neq i}}^{n} \mu_j p_{ji} \min(N_j(t),\, m_j) - \mu_i \min(N_i(t),\, m_i) + \lambda p_{0i}, \quad i = \overline{1,\, n}. \quad (6)$$

It is a linear UDE system with the discontinious right members. It is necessary to solve it by space phase partition into a series of areas and solve each of them. The

system (6) can be solved, for example, by the means of computer mathematics Maple 8 system.

Let's introduce a sign $v_i(t) = M\{V_i(t)\}$, $i = \overline{1, n}$. Then, from (1), (4) we receive passing to a limit at $\Delta t \to 0$, we will receive inhomogeneous linear UDE of the first order

$$\frac{dv_i(t)}{dt} = -\mu_i \min(N_i(t), m_i)(p_{i0}b_{i0} + \sum_{j=1}^{n} p_{ij}b_{ij}) + \\ + \sum_{\substack{j=1 \\ j \neq i}}^{n} \mu_j \min(N_j(t), m_j)p_{ji}a_{ji} + \lambda p_{0i}a_{0i} + c_i, \quad i = \overline{1, n}. \quad (7)$$

Having set entry conditions $v_i(0) = v_{i0}$, $i = \overline{1, n}$, it is possible to discover expected incomes of network systems (average magnitudes of squares occupied in warehouses).

Knowing the expressions for $N_i(t)$ and the average warehouse squares occupied with cargoes, it is possible to design the warehouse squares of subjects S_i, $i = \overline{1, n}$. Let's consider the following modeling example.

Example 1. We will consider the closed network presented on fig. 1, consisting of $n = 15$ one-linear QS, where $K = 70$ is the number of requests in the network. The requests service intensities in the network system lines are equal to: $\mu_1 = \mu_7 = \mu_{12} = 3.5$, $\mu_2 = \mu_5 = \mu_{11} = 2.8$, $\mu_3 = \mu_6 = \mu_{10} = \mu_{14} = 2.1$, $\mu_4 = \mu_8 = 2.2$, $\mu_9 = \mu_{15} = 4.1$, $\mu_{13} = 2.9$ and probabilities of request transitions between QS networks is $p_{15\,i} = \frac{1}{14}$, $p_{i\,15} = 1$, $i = \overline{1, 14}$, let's define also $p_{ii} = -1$, $i = \overline{1, 15}$, remaining $p_{ij} = 0$, $i, j = \overline{1, 15}$.

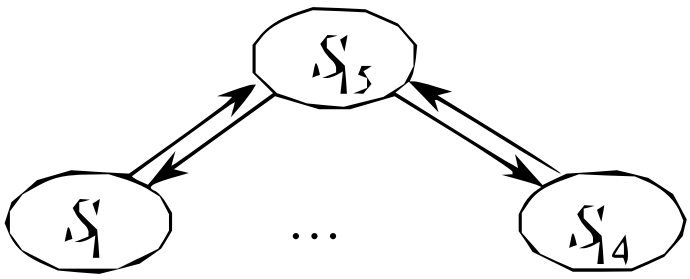

Fig. 1. The network scheme for example 1

Let's assume that the network functions so that on the average there are no queues observed in the peripheral QS, and the central QS functions in the conditions high workload. Then system DDE for the average number of requests in the network systems (5) will be represented as follows:

$$\frac{dN_i(t)}{dt} = \sum_{j=1}^{14} \mu_j p_{ji} N_j(t) + \mu_{15} p_{15i}, \quad i = \overline{1, 15}. \quad (8)$$

Let's set values of income expectations from transitions between network states in the following way:

$$c_i = 9\sin\frac{\pi}{5(i-1)}, \quad i = \overline{1,\,16},$$
$$a_{15\,i} = 1.6, \; i = \overline{1,\,6}, \quad a_{15\,i} = 1.1, \; i = \overline{7,\,13}, \quad a_{15\,14} = 1.35,$$
$$a_{i\,15} = (7,\,11,\,6,\,19,\,28,\,31,\,8,\,17,\,24,\,9,\,12,\,17,\,8,\,13,\,9), \quad i = \overline{1,\,14}.$$

Then expected incomes of network systems discovered by the means of the package Mathematica 5.1 provided that at the initial instant of time $v_i(0) = 5$, $i = \overline{1,\,14}$, $v_{15}(0) = 50$, and entry conditions $N_i(0) = 5$, $i = 1, 7, 9$, $N_i(0) = 3$, $i = 2,\,3,\,12$, $N_i(0) = 4$, $i = 4,\,13,\,14$, $N_i(0) = 6$, $i = 5,\,8,\,11$, $N_i(0) = 2$, $i = 6,\,10$, $N_{15}(0) = 18$, behave as it is shown in fig. 2.

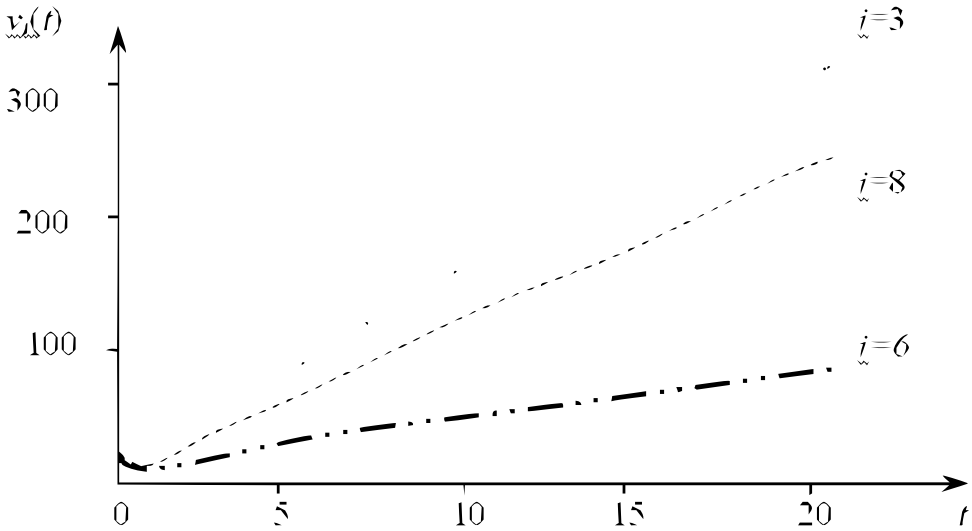

Fig. 2. Expected incomes of systems S_3, S_6, S_8

REFERENCES

1. *Nerush JU.M.* Commercial logistics – Moscow: Banks and stock exchanges, JUNITI. 1997. P. 271.

СИСТЕМА ОБСЛУЖИВАНИЯ С ПОВТОРНЫМИ ВЫЗОВАМИ И ПРИОРИТЕТАМИ

В.И. Клименок[1], Г.К. Мишкой[2]

[1] *Белорусский государственный университет*

[2] *Институт математики и информатики АН Молдовы*

[1] *Минск, Беларусь*

[2] *Кишинёв, Молдова*

[1] klimenok@bsu.by

В статье рассматривается однолинейная система массового обслуживания с неоднородными запросами двух типов, поступающих в систему в $MMAP$ (*Marked Markovian Arrival Process*)-потоке. В случае недоступности обслуживающего прибора запросы 1-го типа идут на орбиту бесконечной емкости (орбита 1), а запросы 2-го типа - на орбиту конечного объема (орбита 2). Находясь на орбитах, запросы делают повторные попытки попасть на обслуживание через случайные интервалы времени. Предполагается, что запросы 2-го типа имеют абсолютный приоритет, что означает следующее: если в момент поступления приоритетного запроса (первичного или повторного) прибор занят обслуживанием неприоритетного запроса, то его обслуживание прерывается и этот запрос идет на орбиту 1, а приоритетный запрос поступает на обслуживание.

Данная работа была частично поддержана Белорусским республиканским фондом фундаментальных исследований через грант № Ф10МЛД-010.

Ключевые слова: система массового обслуживания, $MMAP$-поток, повторные вызовы, две орбиты, абсолютный приоритет.

1. ВВЕДЕНИЕ

Системы массового обслуживания с повторными вызовами широко используются для моделирования процессов передачи и обработки информации в современных телекоммуникационных сетях см., например, [1]. Как правило, системы обслуживания с повторными вызовами рассматриваются в предположении, что в них поступают однородные запросы, которые, в случае занятого обслуживающего прибора (приборов), идут на единственную орбиту. Вместе с тем, типичной для приложений является ситуация, когда на обслуживание поступают запросы разных типов, отличающихся как интенсивностью поступления, так и интенсивностью обслуживания. В случае систем с повторными вызовами запросы разных типов могут отличаться также интенсивностью повторных попыток и нетерпеливостью.

В качестве недавних публикаций по СМО с повторными вызовами и двумя типами запросов можно указать статью испанских авторов [4] и ссылки в ней. Мо-

тивацией для рассмотрения модели, которая является предметом исследования в данной статье, является следущее: (а) неоднородные запросы разных приоритетов могут поступать в систему в коррелированном потоке и, в случае отсутствия свободных приборов, направляться на орбиты различного объема, причем орбита для приоритетных запросов может иметь конечный объем, что косвенно позволяет учесть нетерпеливость запросов из приоритетного потока; (б) как первичные так и повторные приоритетные запросы могут иметь абсолютный приоритет перед запросами из неприоритетного потока.

Приоритетные системы широко представлены в литературе, см., например, известные монографии [2], [3]. Однако подавляющее число публикаций посвящено системам с ожиданием, в которые запросы разных приоритетов поступают в стационарных пуассоновских потоках. В рассматриваемой модели СМО предполагается, что запросы двух приоритетов поступают в систему в соответствии с Marked Markovian Arrival Process ($MMAP$) (маркированный марковский входной поток), см., например, [6].

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается однолинейная система массового обслуживания с неоднородными запросами двух приоритетов, котрые поступают в систему в соответствии с ($MMAP$) (*Marked Markovian Arrival Process* – маркированный марковский входной поток), см., например, [6]). Поступление запросов в таком потоке может происходить только в моменты скачков некоторой неприводимой цепи Маркова $\nu_t,\ t \geq 0$, с непрерывным временем и конечным пространством состояний $\{0, 1, ..., W\}$, которая называется управляющим процессом $MMAP$. Интенсивности переходов процесса $\nu_t,\ t \geq 0$, сопровождающиеся поступлением запроса r-го типа задаются элементами квадратной матрицы D_r, $r = 1, 2$, порядка $W + 1$, а интенсивности "холостых"переходов – недиагональными элементами матрицы D_0. При этом матрица $D(1) = D_0 + D_1 + D_2$ является генератором процесса $\nu_t, t \geq 0$. Средняя интенсивность поступления запросов определяется как $\lambda = \boldsymbol{\theta}(D_1 + D_2)\mathbf{e}$, где $\boldsymbol{\theta}$ – вектор стационарного распределения цепи Маркова $\nu_t, t \geq 0$ – является единственным решением системы $\boldsymbol{\theta} D(1) = \mathbf{0},\ \boldsymbol{\theta}\mathbf{e} = 1$. Здесь и далее $\mathbf{e}$ – вектор-столбец, состоящий из единиц, а $\mathbf{0}$ – вектор-строка, состоящая из нулей. Средняя интенсивность λ_r поступления запросов r-го типа определяется как $\lambda_r = \boldsymbol{\theta} D_r \mathbf{e}$, $r = 1,\ 2$.

Заметим, что $MMAP$ – это обобщение хорошо известного MAP (*Markov Arrival Process*) потока на случай неоднородных запросов. Обзор работ, посвященных системам с повторными вызовами и MAP-потоком, приводится в [5].

Замечание 1. В рамках рассматриваемой модели будем считать неприоритетные и приоритетные запросы запросами 1-го типа и 2-го типа соответственно. Интенсивности поступления неприоритетных и приоритетных запросов задаются матрицами D_1 и D_2 соответственно.

Если прибор свободен в момент поступления любого запроса, этот запрос мгновенно идет на обслуживание. Время обслуживания запроса типа r распределено

по экспоненциальному закону с параметром μ_r, $r = 1, 2$. После окончания обслуживания запрос покидает систему.

Если в момент поступления неприоритетного запроса прибор занят, то поступивший запрос идет на бесконечную орбиту (орбита 1), откуда делает попытки вновь попасть на обслуживания через случайные интервалы времени. Если на орбите 1 в некоторый момент времени находится i запросов, то вероятность повторной попытки в интервале $(t, t + \Delta t)$ равна $\alpha_i \Delta t + o(\Delta t)$, где $\alpha_i \to \infty$ при $i \to \infty$, $\alpha_0 = 0$.

Если в момент поступления приоритетного запроса прибор занят обслуживанием неприоритетного запроса, то его обслуживание прерывается и этот запрос идет на орбиту 1, а приоритетный запрос поступает на обслуживание. Если в момент поступления приоритетного запроса прибор занят обслуживанием приоритетного же запроса, то поступивший запрос идет на конечную орбиту (орбита 2) размерности R , откуда делает попытки попасть на обслуживание через экспоненциально распределенное время с параметром γ. Если же орбита 2 заполнена, то поступивший запрос теряется.

Если в момент повторной попытки с орбиты 2 прибор занят обслуживанием неприоритетного запроса, то его обслуживание прерывается и этот запрос идет на орбиту 1, а приоритетный запрос поступает на обслуживание.

3. ПРОЦЕСС, ОПИСЫВАЮЩИЙ ФУНКЦИОНИРОВАНИЕ СИСТЕМЫ

Поведение рассматриваемой СМО определяется четырехмерной цепью Маркова

$$\xi_t = \{i_t, n_t, r_t, \nu_t, \}, \ t \geq 0,$$

где $i_t, \ i_t \geq 0$, – число запросов на первой орбите; $r_t, \ r_t = \overline{1, R}$, – число запросов на второй орбите; n_t принимает значения 0,1,2 в зависимости от того, свободен прибор в момент времени t, обслуживает запрос первого типа либо обслуживает запрос второго типа, ν_t – состояние $MMAP$-потока, $\nu_t = \overline{0, W}$, в момент времени t.

Цепь Маркова $\xi_t, t \geq 0$, имеет одну счетную компоненту i_t и три конечных компоненты n_t, r_t и ν_t. Перенумеруем состояния цепи в лексикографическом порядке и будем называть множество состояний имеющих значение i компоненты i_t *уровнем* i. Пусть $Q_{i,l}$, $i, l \geq 0$, – матрица интенсивностей переходов из состояний, включенных в уровень i, в состояния уровня l.

Введем следующие обозначения:

$$\bar{R} = R + 1, \ \bar{W} = W + 1, \ K = (R + 1)(W + 1), \Gamma = \gamma diag\{0, 1, \ldots, R\}.$$

Лемма 1. *Инфинитезимальный генератор Q цепи Маркова $\xi_t = \{i_t, n_t, r_t, \nu_t, \}$, $t \geq 0$, имеет следующую блочную структуру:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & O & \dots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & O & \dots \\ O & Q_{2,1} & Q_{2,2} & Q_{2,3} & \dots \\ O & O & Q_{3,2} & Q_{3,3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

где ненулевые блоки порядка $3K \times 3K$ имеют следующий вид:

$$Q_{i,i-1} = \begin{pmatrix} O & \alpha_i I & O \\ O & O & O \\ O & O & O \end{pmatrix}, \ i \geq 1, \quad Q_{i,i+1} = \begin{pmatrix} O & O & O \\ O & I_{\bar{R}} \otimes D_1 & \Gamma \oplus D_2 \\ O & O & I_{\bar{R}} \otimes D_1 \end{pmatrix}, \ i \geq 0,$$

$$Q_{i,i} = \begin{pmatrix} Q_{i,i}^{(0,0)} & Q_{i,i}^{(0,1)} & Q_{i,i}^{(0,2)} \\ Q_{i,i}^{(1,0)} & Q_{i,i}^{(1,1)} & O \\ Q_{i,i}^{(2,0)} & O & Q_{i,i}^{(2,2)} \end{pmatrix}, \quad i \geq 0,$$

а матрицы $Q_{i,i}^{(m,n)}$ порядка $K \times K$ определяются как:

$$Q_{i,i}^{(0,0)} = D_0 \oplus (-\Gamma) - \alpha_i I, \ \ Q_{i,i}^{(0,1)} = I_{\bar{R}} \otimes D_1, \quad Q_{i,i}^{(0,2)} = (\Gamma \check{I}) \oplus D_2,$$

$$Q_{i,i}^{(1,0)} = \mu_1 I, \ \ Q_{i,i}^{(1,1)} = (-\Gamma) \oplus D_0 - \mu_1 I,$$

$$Q_{i,i}^{(2,0)} = \mu_2 I, \ \ Q_{i,i}^{(2,2)} = I_{\bar{R}} \otimes D_0 + \hat{I} \otimes D_2 - \mu_2 I.$$

Здесь I и O – единичная и нулевая матрицы , $\otimes$ – символ кронекерова произведения матриц, $\hat{I}$ и $\check{I}$ - квадратные матрицы порядка $\bar{R}$,

$$\hat{I} = \begin{pmatrix} 0 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \ \check{I} = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Следствие 1. *Цепь Маркова $\xi_t, t \geq 0$, принадлежит классу многомерных асимптотически квазитеплицевых цепей Маркова (АКТЦМ), см. [7].*

Носителем асиптотических свойств АКЦМ является вложенная предельная квазитеплицева цепь Маркова с дискретным временем, матрицы вероятностей переходов которой определяются следующим образом:

$$P_{i,i-1} = Y_0 = \lim_{i\to\infty} R_i^{-1} Q_{i,i-1}, \ P_{i,i} = Y_1 = \lim_{i\to\infty} R_i^{-1} Q_{i,i} + I, \ \ P_{i,i+1} = Y_2 = \lim_{i\to\infty} R_i^{-1} Q_{i,i+1},$$

где R_i – диагональная матрица, имеющая положительные диагональные элементы, которые с точностью до знака совпадают с соответствующими диагональными элементами матрицы $Q_{i,i}$.

Следствие 2. *Матрицы вероятностей переходов вложенной предельной цепи Маркова имеют следующий вид:*

$$Y_0 = \begin{pmatrix} O & I & O \\ O & O & O \\ O & O & O \end{pmatrix}, \quad Y_2 = \begin{pmatrix} O & O & O \\ O & R_1^{-1}(I \otimes D_1) & R_1^{-1}(\Gamma \oplus D_2) \\ O & O & R_2^{-1}(I \otimes D_1) \end{pmatrix},$$

$$Y_1 = \begin{pmatrix} O & O & O \\ Y_1^{(1,0)} & Y_1^{(1,1)} & O \\ Y_1^{(2,0)} & O & Y_1^{(2,2)} \end{pmatrix},$$

где

$$Y_1^{(1,0)} = R_1^{-1}\mu_1, \;\; Y_1^{(1,1)} = R_1^{-1}[(-\Gamma) \oplus (D_0 - \mu_1 I)] + I,$$
$$Y_1^{(2,0)} = R_2^{-1}\mu_2, \;\; Y_1^{(2,2)} = R_2^{-1}[I \otimes (D_0 - \mu_2 I) + \hat{I} \otimes D_2] + I.$$

4. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ. ХАРАКТЕРИСТИКИ ПРОИЗВОДИТЕЛЬНОСТИ

Теорема 1. *Для того, чтобы цепь Маркова $\xi_t, \;\; t \geq 0$, была эргодической, достаточно выполнения неравенства:*

$$\frac{\lambda_1}{\mu_1}(1 + \frac{\lambda_2 + \gamma R}{\mu_2}) < 1. \tag{1}$$

Далее предполагаем, что неравенство (1) выполняется. Тогда стационарные вероятности

$$\pi(i, r, n, \nu) = \lim_{t \to \infty} P\{i_t = i, r_t = r, n_t = n, \nu_t = \nu\}, \; i \geq 0, r = \overline{0, R}, n = 0, 1, 2, \nu = \overline{0, W},$$

существуют. Скомбинируем вероятности $\pi(i, r, n, \nu)$ состояний, входящих в уровень i, в вектор-строки $\boldsymbol{\pi}_i, \; i \geq 0$. Для вычисления этих векторов используется алгоритм для вычисления стационарного распределения АКТЦМ, разработанный в [7].

Найдя векторы $\boldsymbol{\pi}_i, \; i \geq 0$, можно вычислить различные характеристики производительности исследуемой системы. Наиболее важные из них приведены ниже.

- Среднее число запросов на орбитах (орбите 1 и орбите 2)

$$L_1 = \sum_{i=1}^{\infty} i \boldsymbol{\pi}_i \mathbf{e}, \; L_2 = \sum_{i=0}^{\infty} \boldsymbol{\pi}_i (\mathbf{e}_3 \otimes diag\{0, 1, \ldots, R\} \otimes \mathbf{e}_{\bar{W}}) \mathbf{e}.$$

- Вероятность того, что прибор свободен $\quad P^{(0)} = \sum\limits_{i=0}^{\infty} \boldsymbol{\pi}_i (diag\{1, 0, 0\} \otimes I_{\bar{R}\bar{W}}) \mathbf{e}.$
- Вероятность того, что прибор обслуживает неприоритетный запрос

$$P_{serv}^{(1)} = \sum_{i=0}^{\infty} \boldsymbol{\pi}_i (diag\{0, 1, 0\} \otimes I_{\bar{R}\bar{W}}) \mathbf{e}.$$

- Вероятность того, что прибор обслуживает приоритетный запрос

$$P_{serv}^{(2)} = \sum_{i=0}^{\infty} \boldsymbol{\pi}_i (diag\{0,0,1\} \otimes I_{\bar{R}\bar{W}})\mathbf{e}.$$

- Вероятность потери приоритетного запроса

$$P_{loss} = {\lambda_2}^{-1} \sum_{i=0}^{\infty} \boldsymbol{\pi}_i (diag\{0,1,1\} \otimes diag\{0,0,\ldots,0,1\} \otimes D_2)\mathbf{e}.$$

- Вероятность того, что неприоритетный запрос начнет обслуживаться сразу после поступления

$$P_{imm}^{(1)} = {\lambda_1}^{-1} \sum_{i=0}^{\infty} \boldsymbol{\pi}_i (diag\{1,0,0\} \otimes I_{\bar{R}} \otimes D_1)\mathbf{e}.$$

- Вероятность того, что приоритетный запрос начнет обслуживаться сразу после поступления

$$P_{imm}^{(1)} = {\lambda_2}^{-1} \sum_{i=0}^{\infty} \boldsymbol{\pi}_i (diag\{1,1,0\} \otimes I_{\bar{R}} \otimes D_2)\mathbf{e}.$$

- Вероятность того, что произвольный неприоритетный запрос успешно обслужится (не посетив орбиту)

$$P_{succ}^{(1)} = {\lambda_1}^{-1}\mu_1 \sum_{i=0}^{\infty} \boldsymbol{\pi}_i [(1,0,0)^T \otimes diag\{D_1((r\gamma+\mu_1)I - D_0 - D_1)^{-1},\ r = \overline{0,R}\}]\mathbf{e}.$$

- Вероятность того, что произвольный неприоритетный запрос хотя бы раз посетит орбиту $P_{orb} = 1 - P_{succ}^{(1)}$.

5. ЗАКЛЮЧЕНИЕ

Рассмотрена однолинейная система обслуживания с абсолютными приоритетами и повторными вызовами. Запросы разных приоритетов поступают в систему в $MMAP$-потоке, имеют разные орбиты, различные интенсивности обслуживания и повторов. Поведение системы описывается четырехмерной асимптотически квазитеплицевой цепью Маркова с непрерывным временем. Получено достаточное условие существования стационарного режима в системе, формулы для основных характеристик ее производительности, включая математические ожидания числа повторных вызовов на орбитах, вероятность потери приоритетного запроса, вероятность немедленного обслуживания запросов разных приоритетов.

ЛИТЕРАТУРА

1. Artalejo, J.R. and Comez-Corral, A.: Retrial queueing systems: a computational approach. Springer, Berlin-Heidelberg (2008)

2. В.В. Гнеденко, Э.А. Даниелян, Г.П. Климов, В.Ф. Матвеев. Приоритетные системы обслуживания. М.: МГУ.1973.
3. Матвеев В.Ф., Ушаков В.Г. Системы массового обслуживания. М.: Изд-во МГУ, 1984. 240 с.
4. Domenech-Benlloch M.J., Gimenez-Guzman J.M., Pla V., Casarea-Giner V., Martinez-Bauset J.: Solving multiserver systems with two retrial orbits using value extrapolation: a comparative perspective. In: Al-Begain et al (eds.)Analytical and Stochastic Modelling Techniques and Applications. LNCS, vol. 5513, pp. 56-70. Springer, Heidelberg (2009)
5. Gomez-Corral A.: A bibliographical guide to the analysis of retrial queues through matrix analytic techniques. Annals of Operations Research. 141, 163–191 (2006)
6. He Q.M.: Queues with marked customers. Advances in Applied Probability. 28, 567-587 (1996)
7. Klimenok V.I., Dudin A.N.: Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory. Queueing Systems. 54, 245–259 (2006)

ОБ ИССЛЕДОВАНИИ ВХОДНЫХ ПОТОКОВ СООБЩЕНИЙ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ МЕЖБАНКОВСКИХ РАСЧЕТОВ

Е. Колузаева[1*], М. Кищенко[2], М. Маталыцкий[1]

[1] *Гродненский государственный университет имени Янки Купалы,*

[2] *Расчетный центр Национального банка Республики Беларусь*

[1] *Гродно, Беларусь*

[2] *Минск, Беларусь*

[*] koluzaeva@gmail.com

В статье исследуются входные потоки электронных сообщений, поступающих в автоматизированную систему межбанковских расчетов. Описана методика нахождения законов распределения интервалов времен между поступлением электронных сообщений, приведены примеры анализа входных потоков сообщений от двух банков. Результаты исследований могут быть использованы для выбора типов систем массового обслуживания с целью их применения при решении задач моделирования фрагментов указанной автоматизированной системы в виде сети массового обслуживания.

Ключевые слова: автоматизированная система межбанковских расчетов, входной поток электронных сообщений

1. ВВЕДЕНИЕ

Автоматизированная система межбанковских расчетов (далее - АС МБР) является ключевым компонентом платежной системы Республики Беларусь. Основными функциональными составляющими АС МБР являются система расчетов в режиме реального времени по межбанковским платежам и транспортная система, обеспечивающая надежную и безопасную передачу электронных платежных документов и сообщений (далее - ЭС).

При проектировании и модернизации функционирования АС МБР требуется определить топологию программно-технических средств, необходимые параметры, которые обеспечат достаточную производительность и выполнение других показателей назначения системы. Одним из подходов к решению этих задач является построение модели прохождения потоков ЭС, используемой в качестве инструмента

для отслеживания текущего состояния системы в установленных точках прохождения сообщений, а также для построения краткосрочных и долгосрочных прогнозов по исследуемым показателям. Кроме того, разработанные модели могут применяться для решения задач оптимизации параметров технологической обработки сообщений в АС МБР.

На первом этапе построения модели осуществляется исследование входных потоков ЭС, поступающих от банков. Опишем методику исследования[1], приведем примеры по данным двух банков, от программно-технических комплексов (далее - ПТК) которых поступают сообщения в АС МБР. В примерах для краткости входные потоки сообщений от ПТК двух банков представлены как входные потоки ЭС от ПТК 1 и ПТК 2 соответственно (нумерация условная, используется при исследованиях).

2. АНАЛИЗ ВХОДНЫХ ПОТОКОВ

2.1. Методика исследования входных потоков. Для исследования входных потоков ЭС рассмотрим выборки интервалов времен между последовательными поступлениями ЭС, измеренные, например, с точностью до секунды. Из-за выбранной точности измерения в момент наблюдения может поступать сразу несколько ЭС. Рассматриваемые интервалы времен являются случайной величиной τ, распределенной по некоторому закону, который следует определить. Для нахождения этого закона требуется построить выборки значений случайной величины τ по данным, измеренным за каждый рабочий день некоторого месяца.

Методику нахождения закона распределения случайной величины τ опишем более подробно. По данным входного потока ЭС за некоторый день рассматривается выборка значений случайной величины τ объема N, по которой строится вариационный ряд

$$\begin{array}{c|c|c|c|c|} \tau_i & \tau_1 & \tau_2 & \dots & \tau_k \\ \hline n_i & n_1 & n_2 & \dots & n_k \end{array}, \tag{1}$$

где τ_i – i-е значение случайной величины τ, $\tau_1 < \tau_2 < \ldots < \tau_k$, с выборочными частотами n_1, n_2,..., n_k, $N = n_1 + n_2 + ... + n_k$.

Далее, используя ряд (1), строится согласно определению эмпирическая функция распределения $F_e(t)$ по формуле

$$F_e(t) = \sum_{i=1}^{k} P\{\tau_i : \tau_i < t\} = \sum_{\{i:\, \tau_i < t\}} \frac{n_i}{N} \tag{2}$$

и рассчитывается выборочное среднее по формуле

[1]Исследования проводятся в рамках выполнения работ по договору № 22/D о научно-техническом сотрудничестве между Гродненским государственным университетом имени Янки Купалы и Национальным банком Республики Беларусь от 29.01.2010

$$\bar{\tau} = \frac{1}{N}\sum_{i=1}^{k} \tau_i n_i. \qquad (3)$$

По виду графика эмпирической функции распределения определяется характер распределения случайной величины τ и предполагается вид теоретического закона распределения $F_\tau(t)$ с m параметрами.

Для теоретической функции распределения $F_\tau(t)$ проверяется гипотеза H_0, которая заключается в том, что $F_\tau(t) = F_e(t)$, при конкурирующей гипотезе H_1: $F_\tau(t) \neq F_e(t)$.

Для проверки гипотезы H_0, согласно правилу Стерджеса [1, 2], разбивается диапазон наблюдаемых значений $[\tau_{\min}, \tau_{\max}]$ на $k = round(3.32 \lg N + 1)$ непересекающихся промежутков $[z_{i-1}, z_i)$, $i = \overline{1, k}$, где $z_0 < z_1 < \ldots < z_k$ – точки, соответствующие границам интервалов, причем $z_0 = \tau_{\min} - \frac{h}{2}$, $z_k = \tau_{\max} + \frac{h}{2}$, $h = \frac{\tau_{\max} - \tau_{\min}}{k}$ (функция $round(x)$ округляет число x до ближайшего целого значения).

Для каждого i-го интервала находим эмпирические частоты $\nu_i = \sum\limits_{\{z_{i-1} \le \tau_i < z_i\}} n_i$ и теоретические вероятности $p_i = F_\tau(z_i) - F_\tau(z_{i-1})$, где $p_i > 0$ – теоретическая вероятность попадания значений случайной величины τ в i-й интервал. Далее строим статистику γ_N по формуле

$$\gamma_N = \sum_{i=1}^{k} \frac{(\nu_i - Np_i)^2}{Np_i} \ge 0. \qquad (4)$$

При условии, что гипотеза H_0 верна, при $N \to \infty$ статистика γ_N имеет распределение χ^2 с $(k-m-1)$ степенями свободы [2] и называется *хи-квадрат статистикой*.

Гипотеза H_0 принимается, если $\gamma_N \le \Delta$ и отклоняется, если $\gamma_N > \Delta$, где пороговое значение $\Delta = \Delta(\varepsilon) = F_\gamma^{-1}(1-\varepsilon)$ – квантиль уровня $(1-\varepsilon)$ распределения $F_\gamma(u)$ [3]. Таким образом, если $\gamma_N \le \Delta(\varepsilon)$, то с вероятностью $(1-\varepsilon)$ верна гипотеза H_0.

Так как на интервале наблюдения из-за выбранной точности измерения (1с) входного потока от банка-отправителя может поступать в АС МБР сразу несколько ЭС, то нужно дополнительно найти закон распределения случайной величины η, значения которой равны числу ЭС, поступающих за интервал наблюдения в одну секунду. Для этого по полученной выборке объема N, аналогично как и для случайной величины τ, строится вариационный ряд и вычисляются вероятности $P_i = P\{\eta = \eta_i\} = \frac{l_i}{N}$, где l_i – частота появления значения η_i в выборке, рассчитывается выборочное среднее $\bar{\eta} = \sum\limits_{i=1}^{k} \eta_i P_i$.

2.2. Исследование входного потока ЭС от ПТК 1. По данным за некоторый день, измеренным с точностью до секунды, была получена выборка значений случайной величины τ объема $N = 176$ – число наблюдаемых значений. Для этой выборки был построен вариационный ряд (1), найдено выборочное среднее $\bar{\tau} =$

163,914 с и построена по формуле (2) эмпирическая функция распределения, график которой изображен на рис.1. Для построения графика использовался пакет эконометрического моделирования EViews 3.1.

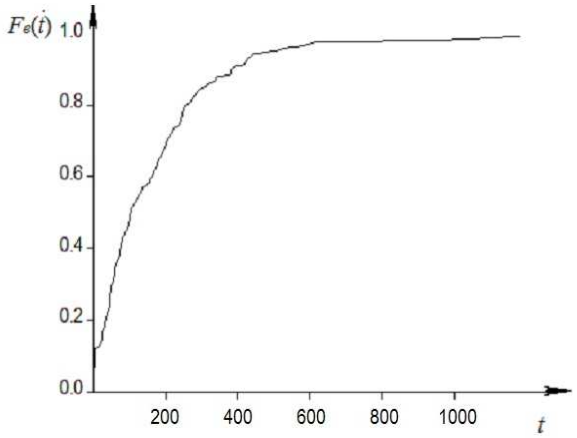

Рис. 1. График эмпирической функции распределения $F_e(t)$

По графику виден экспоненциальный характер эмпирической функции распределения, поэтому в качестве теоретического закона распределения предполагаем экспоненциальный закон, функция распределения которого равна

$$F_\tau(t) = 1 - e^{-\lambda t}, \tag{5}$$

где λ – параметр экспоненциального распределения, $\lambda = \dfrac{1}{\bar{\tau}} = \dfrac{1}{163,914} = 6,1\times$ $\times 10^{-3}$ (1/с) – интенсивность поступления.

Для построенной выборки минимальное значение равно $\tau_{\min}$ =1с, максимальное – $\tau_{\max} = 1\ 179$ с, а число интервалов разбиения наблюдаемых значений $k = 8$ шириной $h = 147{,}25$ с. Для каждого интервала были найдены эмпирические частоты ν_i и теоретические вероятности p_i, $i = \overline{1,\ 8}$. Так как в качестве теоретической функции распределения была взята экспоненциальная функция распределения, то $p_i = F_\tau(z_i) - F_\tau(z_{i-1}) = e^{-\lambda z_{i-1}} - e^{-\lambda z_i}$.

Для построенного интервального ряда при проверке гипотезы о законе распределения значение статистики γ_N, построенной по формуле (4) для уровня значимости $\varepsilon = 0,05$, получилось равным 1,095, а ее критическое значение равно $\Delta(\varepsilon) = 7,815$. Таким образом, так как $\gamma_N \leq \Delta(\varepsilon)$, то с вероятностью 0,95 можно принять гипотезу об экспоненциальном законе распределения случайной величины τ.

Так как за интервал наблюдения в одну секунду может поступать сразу несколько ЭС, то был найден закон распределения случайной величины η, значения которой равны числу ЭС, поступающих за интервал наблюдения (1с). Для η была получена выборка объема $N = 176$, построен вариационный ряд и найдены вероятности P_i, $i = \overline{1,31}$, вычислено выборочное среднее $\bar{\eta} = 11,528$ ЭС.

2.3. Исследование входного потока ЭС от ПТК 2. По данным за некоторый день, измеренным с точностью до секунды, была получена выборка значений случайной величины τ объема $N = 472$. Для проверки гипотезы о законе распределения случайной величины τ для исследуемого вариационного ряда были найдены выборочное среднее $\bar{\tau} = 70,62$ с, максимальное и минимальное значения случайной величины: $\tau_{\min} = 1$с, $\tau_{\max} = 3\ 409$ с, количество интервалов разбиения $k = 10$ и ширина интервала разбиения $h = 34{,}08$ с.

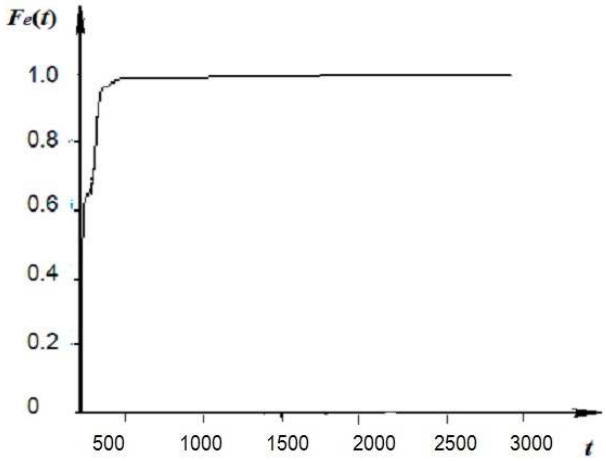

Рис. 2. График эмпирической функции распределения $F_e(t)$

Для каждого интервала найдены эмпирические частоты ν_i и теоретические вероятности p_i, построена эмпирическая функция распределения, рис. 2, по ее виду нельзя сделать однозначный вывод о законе распределения случайной величины τ, поэтому проверялось несколько гипотез: об экспоненциальном законе распределения, о гамма-распределении, о логарифмически нормальном распределении и распределении Вейбулла, так как графики функций распределений этих законов имеют схожий вид.

Проверим гипотезу о том, что случайная величина τ распределена по закону Вейбулла

$$F_\tau(x) = 1 - e^{-(x\Theta)^\alpha}.$$

Теоретические вероятности находятся по формуле $p_i = F_\tau(z_i) - F_\tau(z_{i-1}) = e^{-(z_{i-1}\Theta)^\alpha} - e^{-(z_i\Theta)^\alpha}$. Поскольку число параметров в распределении Вейбулла равно двум, то пороговое значение хи-квадрат статистики для уровня значимости $\varepsilon = 0,05$ с $(k–3)$ степенями свободы, рассчитанное по формуле (4), равно $\gamma_N = 6{,}456 < \Delta(0,05) = 7{,}815$. Значит можно принять гипотезу о том, что случайная величина τ распределена по закону Вейбулла с параметрами $\alpha = 0{,}2$, $\Theta = \dfrac{\bar{\tau}}{\Gamma\left(1 + \frac{1}{\alpha}\right)} = 58{,}835$. Значение параметра α находилось из условия минимума статистики γ_N.

Был также найден закон распределения случайной величины η, значения которой равны числу ЭС, поступающих за интервал наблюдения (1 с). Аналогично как в предыдущем разделе для η была получена выборка объема $N = 472$, построен вариационный ряд и найдены вероятности P_i, $i = \overline{1, 78}$, вычислено выборочное среднее $\bar{\eta} = 37,295$ ЭС.

3. ВЫВОДЫ

Результаты проведенных исследований показали, что интервалы времен между поступающими ЭС от разных банков распределены по различным законам, не всегда совпадающим с экспоненциальным.

Для дальнейшего моделирования прохождения ЭС в АС МБР кроме характеристик входных потоков ЭС следует установить законы обработки ЭС и определить функциональные зависимости технологических параметров прохождения ЭС на различных фазах их обработки в АС МБР. Отметим, что в [4, 5] рассматривались постановки некоторых подобных задач.

ЛИТЕРАТУРА

1. *Закс Л.* Статистическое оценивание. М.: Статистика, 1976.
2. *Харин Ю. С., Малюгин В. И., Харин А. Ю.* Эконометрическое моделирование. Мн.: БГУ, 2003.
3. *Большов Л. Н., Смирнов Н. В.* Таблицы математической статистики. М.: Наука, 1983.
4. *Карпук А. А., Кищенко М. Г.* О математическом моделировании систем межбанковских и межфилиальных расчетов // Технологии информатизации и управления: сб. научн. ст., Мн.: БГУ, 2009. С. 20–24.
5. *Карпук А. А., Маталыцкий М. А.* О математическом моделировании системы межбанковских расчетов // Устойчивое развитие экономики: состояние, проблемы, перспективы: материалы четвертой международной научно-практической конференции, Пинск: ПолесГу, 2010. С. 21–25.

СИСТЕМА MR|M|∞ В УСЛОВИИ РАСТУЩЕГО ВРЕМЕНИ ОБСЛУЖИВАНИЯ

И. Лапатин*, **С. Лопухова**

Томский государственный университет
Томск, Россия
* ilapatin@mail.ru

В работе рассматривается система массового обслуживания с неограниченным числом приборов, на вход которой поступает поток марковского восстановления. Исследуется стационарное распределение числа занятых приборов и выходящий поток системы в условии растущего времени обслуживания.

Ключевые слова: поток марковского восстановления, системы массового обслуживания с неограниченным числом приборов, выходящий поток.

1. ВВЕДЕНИЕ

Системы массового обслуживания [1] с неограниченным числом приборов являются моделями реальных систем в различных сферах повседневной жизни: банковское дело, страхование, транспорт, торговля и т.д.

В случае простейшего входящего потока для таких систем удается найти аналитическое решение для стационарного распределения вероятностей числа занятых приборов в системе, а также для числа заявок, закончивших обслуживание на некотором промежутке времени [2]. Но в случае других моделей входящего потока общих подходов к исследованию нет. Для их исследования применяют методы численного анализа, имитационного моделирования и асимптотического анализа. Анализ числа занятых приборов в системе BMAP|GI|∞ можно найти, например, в работе немецкого ученого Д.Баума [3]. Асимптотический анализ выходящего потока системы MAP|GI|∞ приведен в работе [4].

В данной работе приводится исследование системы массового обслуживания с неограниченным числом приборов и экспоненциальным временем обслуживания, на вход которой поступает поток марковского восстановления. Рассматривается стационарное распределение числа занятых приборов и выходящий поток в условии растущего времени обслуживания.

2. ОПРЕДЕЛЕНИЕ ПОТОКА МАРКОВСКОГО ВОССТАНОВЛЕНИЯ

Рассмотрим двумерный однородный марковский случайный процесс $\{\xi(n);\tau(n)\}$ с дискретным временем $n = 1, 2, 3, ...$, первая компонента $\xi(n)$ которого принима-

ет значения из некоторого дискретного множества. Вторая компонента $\tau(n)$ рассматриваемого процесса принимает неотрицательные значения (вообще говоря, из непрерывного множества).

По определению, марковской переходной функцией $F(k_2, x; k_1, y)$ двумерного однородного марковского случайного процесса $\{\xi(n); \tau(n)\}$ называется

$$F(k_2, x; k_1, y) = P\{\xi(n+1) = k_2, \tau(n+1) < x | \xi(n) = k_1, \tau(n) = y\}$$

Будем рассматривать только такие двумерные случайные процессы $\{\xi(n); \tau(n)\}$, для марковских переходных функций которых выполняются равенства

$$\begin{gathered} F(k_2, x; k_1, y) = \\ = P\{\xi(n+1) = k_2 | \xi(n) = k_1\} \cdot P\{\tau(n+1) < x | \xi(n) = k_1\} = \\ = P_{k_1 k_2} \cdot G_{k_1}(x) \end{gathered} \tag{1}$$

то есть марковская переходная функция не зависит от второго условия (не зависит от значений второй компоненты), а распределение значений процесса $\tau(n+1)$ зависит лишь от значения процесса $\xi(n)$ в предыдущий момент времени.

В силу свойства (1), первая компонента $\xi(n)$ рассматриваемого двумерного марковского процесса $\{\xi(n); \tau(n)\}$ также является марковским процессом, точнее цепью Маркова с дискретным временем и матрицей P вероятностей переходов за один шаг. А функции $G_k(x)$ - это набор условных функций распределения, которые будем записывать в виде диагональной матрицы $G(x)$.

Случайный поток однородных событий

$$t_1 < t_2 < t_3 < ...$$

будем называть потоком марковского восстановления или MR-потоком, заданным матрицей вероятностей перехода P и набором условных функций распределения $G_k(x)$, если для длин $\tau_{n+1} = t_{n+1} - t_n$ его интервалов выполняются равенства

$$\tau_{n+1} = \tau(n)$$

Для дальнейшего исследования определим процесс $k(t) = \xi(n), t_n \leq t < t_{n+1}$

3. ВЫВОД УРАВНЕНИЙ ДЛЯ ХАРАКТЕРИСТИК СИСТЕМЫ MR|M|∞

Рассмотрим систему массового обслуживания с неограниченным числом приборов, на вход которой поступает поток марковского восстановления, заданный матрицей вероятностей переходов P и диагональной матрицей $G(x)$ условных функций распределения. Время обслуживания поступающих заявок будем считать экспоненциальным с параметром μ, одинаковой для всех заявок.

Обозначим $z(t)$ - длину интервала от момента t до момента наступления следующего события во входящем потоке, $i(t)$ - число занятых приборов системы в

момент времени t, $m(t)$ - число заявок, закончивших обслуживание к моменту времени t, а процесс $k(t)$ определен выше.

Для распределения вероятностей

$$P(k,z,i,m,t) = P\{k(t)=k, z(t)<z, i(t)=i, m(t)=m\}$$

четырехмерного марковского процесса $\{k(t), z(t), i(t), m(t)\}$ запишем систему дифференциальных уравнений Колмогорова

$$\begin{gathered}\frac{\partial P(k,z,i,m,t)}{\partial t} = \frac{\partial P(k,z,i,m,t)}{\partial z} - \frac{\partial P(k,0,i,m,t)}{\partial z} + \\ + (i+1)\mu P(k,z,i+1,m-1,t) - i\mu P(k,z,i,m,t) + \\ + \sum_{\nu} \frac{\partial P(\nu,0,i-1,m,t)}{\partial z} G_{\nu}(z) P\end{gathered} \tag{2}$$

Обозначив функции

$$H(k,z,u,x,t) = \sum_i e^{jui} \sum_m e^{jxm} P(k,z,i,m,t),$$

систему (2) для функций $H(k,z,u,x,t)$ запишем в виде

$$\begin{gathered}\frac{\partial H(k,z,u,x,t)}{\partial t} = \frac{\partial H(k,z,u,x,t)}{\partial z} - \frac{\partial H(k,0,u,x,t)}{\partial z} - \\ -j\mu(e^{jx}e^{-ju}-1)\frac{\partial H(k,z,u,x,t)}{\partial u} + \\ +e^{ju}\sum_{\nu} \frac{\partial H(\nu,0,u,x,t)}{\partial z} G_{\nu}(z) P\end{gathered}$$

которую перепишем в матричном виде

$$\begin{gathered}\frac{\partial H(z,u,x,t)}{\partial t} = \frac{\partial H(z,u,x,t)}{\partial z} - \frac{\partial H(0,u,x,t)}{\partial z} - \\ - j\mu(e^{jx}e^{-ju}-1)\frac{\partial H(z,u,x,t)}{\partial u} + \\ +e^{ju}\frac{\partial H(0,u,x,t)}{\partial z} P G(z)\end{gathered} \tag{3}$$

где $H(z,u,x,t) = \{H(1,z,u,x,t), H(2,z,u,x,t), ...\}$.

Дифференциально-матричное уравнение (3) будем решать методом асимптотического анализа в условии растущего времени обслуживания, то есть при $\mu \to 0$.

4. АСИМПТОТИКА РАСТУЩЕГО ВРЕМЕНИ ОБСЛУЖИВАНИЯ

Так как выходящий поток существенно зависит от числа занятых приборов в системе, то сначала будем рассматривать стационарное распределения числа заявок в системе MR|M|∞.

4.1. Исследования числа занятых приборов в системе MR|M|∞. Для исследования стационарного распределения числа занятых приборов в уравнении (3) положим $x = 0$, а $t \to \infty$. Обозначая $h(z,u) = H(z,u,0,\infty)$, получим уравнение

$$\frac{\partial h(z,u)}{\partial z} = \frac{\partial h(0,u)}{\partial z}\left(I - e^{ju}P \cdot G(z)\right) + j\mu\left(e^{-ju} - 1\right)\frac{\partial h(z,u)}{\partial u}. \tag{4}$$

Здесь I - единчная матрица соответствующей размерности.

Полагая, что ε некоторый малый параметр, в уравнении (4) сделаем следующие замены

$$\mu = \varepsilon,\ u = \varepsilon w,\ h(z,u) = F_1(z,w,\varepsilon), \tag{5}$$

тогда для функций $F_1(z,w,\varepsilon)$ получим уравнение

$$\frac{\partial F_1(z,w,\varepsilon)}{\partial z} = \frac{\partial F_1(0,w,\varepsilon)}{\partial z}\left(I - e^{j\varepsilon w}P \cdot G(z)\right) + j\left(e^{-j\varepsilon w} - 1\right)\frac{\partial F_1(z,w,\varepsilon)}{\partial w}. \tag{6}$$

Теорема 1. *Если существует предел решения системы (6)*

$$\lim_{\varepsilon \to 0} F_1(z,w,\varepsilon) = F_1(z,w)$$

то

$$F_1(\infty,w)E = \mathbf{e}^{\{jw\kappa_1\}}$$

где E - единичный вектор-столбец соответствующей размерности, а κ_1 имеет смысл интенсивности входящего MR-потока.

Теорема 1 дает первое асимптотическое приближение характеристической функции числа занятых приборов в системе MR|M|∞ в условии растущего времени обслуживания

$$h(\infty,u)E \approx F_1(\infty,w)E = \mathbf{e}^{\{jw\kappa_1\}=\mathbf{e}^{\{ju\frac{\kappa_1}{\mu}\}}}.$$

Для более точного приближения в задаче (4) сделаем замену

$$h(z,u) = h_2(z,u)\mathbf{e}^{\{ju\frac{\kappa_1}{\mu}\}},$$

тогда получим уравнение

$$\frac{\partial h_2(z,u)}{\partial z} + j\mu\left(1 - e^{-ju}\right)\frac{\partial h_2(z,u)}{\partial u} -$$
$$- \kappa_1\left(1 - e^{-ju}\right)h_2(z,u) + \frac{\partial h_2(0,u)}{\partial z}\left\{e^{ju}P \cdot G(z) - I\right\} = 0 \tag{7}$$

Для выполнения асимптотического исследования в уравнении (7) сделаем следующие замены

$$\mu = \varepsilon^2,\ u = \varepsilon w,\ h_2(z,u) = F_2(z,w,\varepsilon), \tag{8}$$

тогда для функций $F_2(z,w,\varepsilon)$ получим уравнение

$$\frac{\partial F_2(z,w,\varepsilon)}{\partial z} + j\left(1 - e^{-j\varepsilon w}\right)\frac{\partial F_2(z,w,\varepsilon)}{\partial w} -$$
$$- \kappa_1\left(1 - e^{-j\varepsilon w}\right)F_2(z,w,\varepsilon) + \frac{\partial F_2(0,w,\varepsilon)}{\partial z}\left\{e^{j\varepsilon w}P \cdot G(z) - I\right\} = 0 \tag{9}$$

Теорема 2. *Если существует предел решения системы (9)*

$$\lim_{\varepsilon\to 0} F_2(z,w,\varepsilon) = F_2(z,w)$$

то

$$F_2(\infty,w)E = \mathbf{e}^{\left\{\frac{(jw)^2}{2}\kappa_2\right\}},$$

где величина κ_2 определяется равенством

$$\kappa_2 = \kappa_1 + \frac{\partial f_2(0)}{\partial z}E, \tag{10}$$

где вектор $f_2(z)$ является решением уравнения

$$\frac{\partial f_2(z)}{\partial z} + \frac{\partial f_2(0)}{\partial z}(P\cdot G(z) - I) + \frac{\partial R(0)}{\partial z}P\cdot G(z) - \kappa_1 R(z) = 0,$$

здесь $R(z)$ - совместное распределение процессов $k(t)$ и $z(t)$.

Используя результаты теоремы 2 и выражение для функции $h_2(z,u)$ можно записать асимптотическое приближение характеристической функции числа занятых приборов в системе MR|M|∞ в условии растущего времени обслуживания

$$Me^{jui(t)} = h(\infty,u)E \approx \mathbf{e}^{\left\{ju\frac{\kappa_1}{\mu}+\frac{(ju)^2}{2}\frac{\kappa_2}{\mu}\right\}}. \tag{11}$$

4.2. Исследования выходящего потока системы MR|M|∞. Теперь будем исследовать выходящий поток рассматриваемой системы в условии растущего времени обслуживания, для чего в уравнении (3) сделаем следующие замены

$$\mu = \varepsilon,\ u = \varepsilon w,\ H(z,u,x,t) = F(z,w,x,t,\varepsilon),$$

для функций $F(z,w,x,t,\varepsilon)$ получим уравнение

$$\begin{aligned}\frac{\partial F(z,w,x,t,\varepsilon)}{\partial t} &= \frac{\partial F(z,w,x,t,\varepsilon)}{\partial z} - \frac{\partial F(0,w,x,t,\varepsilon)}{\partial z} - \\ &- j(e^{jx}e^{-j\varepsilon w}-1)\frac{\partial F(z,w,x,t,\varepsilon)}{\partial w} + \\ &+ e^{j\varepsilon w}\frac{\partial F(z,w,x,t,\varepsilon)}{\partial z}PG(z).\end{aligned} \tag{12}$$

Теорема 3. *Если существует предел решения системы (12)*

$$\lim_{\varepsilon\to 0} F(z,w,x,t,\varepsilon) = F(z,w,x,t)$$

то

$$F(\infty,0,x,t)E = \mathbf{e}^{\left\{\left(e^{jx}-1\right)\kappa_1 t\right\}}.$$

Теорема 3 дает асимптотическое приближение характеристической функции числа заявок, закончивших обслуживания за некоторое время tв системе MR|M|∞ в условии растущего времени обслуживания

$$Me^{jxm(t)} = H(\infty,0,x,t)E\mathbf{e}^{\left\{\left(e^{jx}-1\right)\kappa_1 t\right\}}. \tag{13}$$

5. ЗАКЛЮЧЕНИЕ

В данной работе была рассмотрена система MR|M|∞. Было получено, что в условии растущего времени обслуживания стационарное распределение числа занятых приборов в системе является асимптотически нормальным (11), а выходящий поток является асимптотически пуассоновским (13). Полученные результаты определяют поведение рассматриваемых характеристик при дочтаточно больших значениях среднего времени обслуживания.

ЛИТЕРАТУРА

1. *Гнеденко Б.В., Коваленко И.Н.* Введение в теорию массового обслуживания. Изд. 3-е, испр. и доп.М.: КомКнига, 2007. 400 с.
2. *Mirasol N.M.* The output of an M|G|∞ queueing system is Poisson // Operations Research. 1963. №. 11. P. 282–284.
3. *Baum. D.* The infinite server queue with Markov additive arrivals in space // Proceedings of the International Conference "Probabilistic analysis of rare events". Riga, Latvia.1999. P. 136–142.
4. *Назаров А.А., Лапатин И.Л.* Асимптотический анализ выходящего потока системы MAP|GI|∞ // Известия политехнического университета. Управление, вычислительная техника и информатика.2009. Т. 315. № 5. С. 191–195.

LIMIT DIFFUSIONS FOR MULTI-CHANNEL NETWORKS WITH INTERDEPENDENT INPUTS

E. Lebedev, A. Chechelnitsky

Kyiv National Taras Shevchenko University
Kyiv, Ukraine
leb@unicyb.kiev.ua

A process of information treatment in multi-channel stochastic networks with interdependent input flows are considered. In heavy traffic conditions functional limit theorems of diffusion approximation types are proved. Local characteristics of the diffusion are represented via network parameters.

Key words: multi-channel network, multi-dimensional Poisson input, diffusion approximation, uniform topology.

Methods of the theory of stochastic processes are an efficient tool for the analysis of the characteristics in the simulation of data transmission networks, computer networks and systems of collective use. This methods permit to evaluate the networks capacity, to find load reserves for network components and control parameters for information flows, as well as, e.g., to control the buffer memory of a node in a packet switching network. Adequate models of the real processes in networks of transmission and processing information are stochastic networks or networks of queues. Their structure is given by probability characteristics of input information flows, processing algorithms and packet-switching schemes.

The process of information treatment which is an object of our interest is a vector of large dimension with complex systems of stochastic relations which determine the process. In previous works, as a rule, input flows were independent. Thus, interdependence of service process components was caused by trajectory intersections for information packets before their outputs from the network. In the paper we omit this restriction and the process of information treatment became more complicated.

That is why the method of functional limit theorems is especially efficient for the analysis of information treatment in stochastic networks under consideration. The method gives the possibility to find those principles which form the foundation of a functioning process for a network of the given type, to construct an approximate process and to calculate the distribution of functionals in order to obtain integral characteristics for the functioning process.

By virtue of conditions imposed on model parameters in our case an approximate process will be diffusion and we will deal with development of the method of diffusion approximation for multi-channel stochastic networks. At first we consider a multi-channel queuing system with multi-dimensional Poisson input flow. The model consists

of r nodes $E_1, \ldots, E_r$. Each node operates as a queuing system $M/M/\propto$. It means that node E_i has a Poisson input flow of calls (exponential interarrival times) $y_i(t)$ with the parameter $\lambda_i > 0, i = 1, 2, \ldots, r$. The calls arrive one at time and immediately take one of servers. Each node has an infinite number of servers. Correspondingly, the service time in the node $E_i, i = 1, 2, \ldots, r$ has exponential distribution with the parameter $\mu_i > 0, i = 1, 2, \ldots, r$. Each call can be served only in one node. After service completion calls leaves the nodes. In addition we suppose that the batch which consists of r calls arrive simultaneously at the all nodes according to a Poisson process $y(t)$ with the parameter $b > 0$. All nodes operate in parallel way in order to provide service in the batch of r calls. The construction of our model implies that in general the calls enter in our model according to a multi -dimensional Poisson input flow $(\nu_1(t), \ldots, \nu_r(t))$ with parameter $\lambda_i > 0, \quad i = 1, 2, \ldots, r, \quad b > 0$, and the component $\nu_i(t)$ has the following representation $\nu_i(t) = y_i(t) + y(t)$.

Such models may be used for planning modern computer networks. It allows to get a high speed and substantial economy of time.

Denote by $Q^{'}(t) = (Q_1(t), \ldots, Q_r(t)), t \in [0, T]$ the number of calls in the nodes of our model at time t. We will study the transient behavior of the multi-dimensional queuing process $Q(t)$ in heavy traffic conditions. This means that characteristics $\mu_i > 0, i = 1, 2, \ldots, r$ of our model depend on a scaling parameter $n, n \to +\infty$ in the following way:

1) $\lim\limits_{n\to\infty} n\mu_i(n) = \mu_i \neq 0, \quad i = 1, 2, \ldots, r.$

Denote by $Q^{(n)}(t) = (Q_1^{(n)}(nt), \ldots, Q_r^{(n)}(nt)), t \in [0, T]$ the number of calls in the nodes of our model at time nt. We assume as well that the initial number of calls is 0 in each node:

2) $Q_i^{(n)}(0) = 0, \quad i = 1, 2, \ldots, r.$

For process $\xi^{(n)}(t) = (\xi_1^{(n)}(t), \ldots, \xi_r^{(n)}(t))$, where $\xi_i^{(n)}(t) = n^{-1/2}(Q_i^{(n)}(nt) - \alpha_i(t)n)$, $\alpha_i(t) = (\lambda_i + b)(1 - e^{-\mu_i t}), i = 1, 2, \ldots, r$ (so $\xi^{(n)}(t)$ is a normalized queuing process $n^{-1/2}(Q^{(n)}(nt) - \alpha(t)n), \alpha(t) = (\alpha_1(t), \ldots, \alpha_r(t))$) we have proved a convergence in uniform topology to a diffusion process.

Theorem 1. *Suppose that conditions 1) and 2) hold. Then the sequence of processes* $\xi^{(n)}(t) = (\xi_1^{(n)}(t), \ldots, \xi_r^{(n)}(t))$ *converges in uniform topology in any interval* $[0, T]$ *to the diffusion process* $\xi(t)$, $\xi(0) = \xi_0 = 0$, *with drift* $a(x) = (a_1(x), \ldots, a_r(x)) = (-\mu_1 x_1, \ldots, -\mu_r x_r)$ *and diffusion matrix*

$$\mathbf{B} = \begin{pmatrix} \lambda_1 + b)(2 - e^{-\mu_1 t}) & b & \ldots & b \\ b & (\lambda_2 + b)(2 - e^{-\mu_2 t}) & b & b \\ \vdots & b & \ddots & b \\ b & \ldots & b & (\lambda_r + b)(2 - e^{-\mu_r t}) \end{pmatrix}$$

Note that the investigation of heavy traffic convergences has a long history and there are several approaches oriented to different classes of queuing models. We prove our theorem using a local approach because it is more convenient for our model. This approach was developed in [1].

The Theorem 1 allows to study instead of functionals of the complex service process (for example, total income connected with service in the network) the corresponding functionals of the diffusion process.

Now we consider more complicated models of multi-channel networks which operate in the following mode.

The network consists of r nodes of information treatment. At the i-th node from the outside information packets arrive in the $\tau_k^{(i)}$, $\quad k = 1, 2, \ldots$, moments of time, $\nu_i(t)$ is the total number of packets arrived in the time interval $[0, t]$. Each of r nodes is a multi-channel system to treat packets of information. If a packet arrives in such system then its treatment begins immediately. For the node i the service time is exponentially distributed with parameter $\mu_i, i = 1, 2, \ldots, r$. The direction of packet motion in the network interior is given by a switching matrix $P = \|p_{ij}\|_1^r$. For any $i = 1, 2, \ldots, r$ $p_{ir+1} = 1 - \sum_{j=1}^r p_{ij}$ is the exit probability for the packet treated in the i-th node. In the used system of notation this network is coded by the symbol $[G|M| \propto]^r$.

As before we will define the process of information treatment in the $[G|M| \propto]^r$-network as an r - dimensional process $Q^{'}(t) = (Q_1(t), \ldots, Q_r(t)), t \geq 0$, where $Q_i(t)$ is the number of information packets in the i-th node at the instant of time t .

Let us study the process of information treatment $Q(t)$ in heavy traffic regime. Heavy traffic regime means: parameters of the network depend on n (the number of series) in such a way that the conditions 1), 2) hold true and the input flow is closed to Brownian motion.

3) There are constants $\lambda_i \geq 0, i = 1, 2, \ldots, r, \quad \lambda_1 + \ldots + \lambda_r \neq 0$, that

$$n^{-1/2}(\nu_1^{(n)}(nt) - \lambda_1 nt, \ldots, \nu_r^{(n)}(nt) - \lambda_r nt) \Rightarrow_{n\to\infty}^{U} W^{'}(t) = (W_1(t), \ldots, W_r(t)),$$

where $W(t)$ is a r-dimensional process of Brownian motion with the null - vector of mean values $EW(1) = 0$ and the correlation matrix $EW(1)W'(1) = \sigma^2 = \|\sigma_{ij}^2\|_1^r$, symbol $\Rightarrow^U$ means weak convergence in uniform topology.

The other parameters of the $[G|M|\infty]^r$ - network do not depend on n.

In the context of conditions 1) - 3) for the open $[G|M|\infty]^r$ - network we shall consider the sequence of stochastic processes

$$\xi^{(n)}(t) = n^{-1/2}(Q^{(n)}(nt) - nq(t)), \quad t \geq 0,$$

where $q^{'}(t) = (q_1(t), \ldots, q_r(t)) = (\theta/\mu)^{'}(I - P(t))$, $(\theta/\mu)^{'} = (\theta_1/\mu_1, \ldots, \theta_r/\mu_r)$, $\theta' = (\theta_1, \ldots, \theta_r) = \lambda'(I - P)^{-1}$ - is a solution of the balance equation for the $[G|M|\infty]^r$ - network, $\lambda' = (\lambda_1, \ldots, \lambda_r)$, $P(t) = \|p_{ij}(t)\|_1^r = exp(Qt)$, $Q = \Delta(\mu)(P - I), \Delta(x) = \|x_i\delta_{ij}\|_1^r$ is a diagonal matrix with the vector on the principal diagonal.

The process $\xi^{(n)}(t)$ may be approximated by diffusion in the following way.

Theorem 2. *Let for a stochastic network of the type $[G|M|\infty]^r$ conditions 1) - 3) be satisfied and the spectral radius of the switching matrix P be strictly less than 1 . Then for any finite interval $[0, T]$ the sequence of stochastic processes $\xi^{(n)}(t), \quad n \geq 1$, converges weakly to the diffusion $\xi(t) \quad (\xi(0) = 0)$ with drift $A(x) = Q^{'}x$ and diffusion matrix $B(t) = \Delta[q^{'}(t)Q] - Q^{'}\Delta[q(t)] - \Delta[q(t)]Q + \sigma^2$ in the uniform topology, as $n \to \infty$.*

A proof of the theorem use a representation of information treatment process as a sum of the indicators on a path of the input flow.

The Theorem 2 extends the results on diffusion approximation of multi-channel networks from [2] by means of restriction removal on input flows.

In summary we make a comment. In our case the switching matrix does not depend on the series parameter n. But if we extend the research field and put

$$P = P_n = P_0 + n^{-1}B_0 + o(n^{-1}),$$

where $P_0 = \|\delta_{\alpha\beta}P^{(\alpha)}\|_1^{r_0}$, $P^{(\alpha)} = \|p_{ij}^{(\alpha)}\|_{i,j\in I_\alpha}$ is an indecomposable stochastic matrix then in the process of approximation it is possible to enlarge the nodes of the initial network: $r \to r_0$, $r_0 \leq r$. The nodes of the set I_α are then joined in a node "α" .

REFERENCES

1. *Gikhman I. I., Skorokhod A. V.* Stochastic differential equations. Kiev: Naukova Dumka. 1982.
2. *Anisimov V. V., Lebedev E. A.* Stochastic networks. Markov models. Kiev: Lybid. 1992.

OPEN NETWORKS WITH MULTIREGIME SERVICE STRATEGIES AND SIGNALS

Y. Letunovich

Gomel State University named after F. Skorina
Gomel, Belarus
yu28031984@yandex.ru

Open queueing networks with several types of customers, Poisson incoming flow, exponential service in the nodes and Markov routing are studied. In each of the nodes there is the only device, which can operate in several regimes. Each regime has a residence time, limited by an exponentially distributed random variable. There are signals, which can increase or reduce the regime of service in the node. The problem of stationary distribution of conditions probabilities form is investigated.

Keywords: queueing networks, multiregime service, reversibility, stationary distribution of conditions probabilities.

1. INTRODUCTION

Queueing networks with multiregime service strategies have been investigated relatively recently. The necessity of their study was caused by practical considerations, because such networks allow us to consider models with partially nonreliable devices. With increase the regime number the regime is getting less reliable and the node productivity decreases.

Transitions from one regime into another are considered as "internal" changes. Under such "internal" changes we mean transitions of the serving device to a less reliable regime due to breakdowns and to more reliable regime because of possible recovery due to natural causes. Moreover, such transitions do not depend solely on the number of the device regime, but also on the types of customers that are in the node.

It is assumed that the time of stay in each regime is limited. After this limited time the device transits with corresponding probabilities to either a regime with a larger number or a regime with a lower number. Described transitions are caused by the properties of each of the regimes and do not depend on the customers in the node.

Considered network is modified by the addition of information signals, which can increase or reduce the regime of service in the node. Presence of signals can be interpreted as an external influence on the network.

2. MODEL DESCRIPTION

We consider open queueing network with M types of customers, which contains N nodes. There are three Poisson input flows: the flow of customers with parameter λ

and two flows of signals, which can reduce or increase the number of regime. They have rates ω^- and ω^+ accordingly.

Every customer of input flow passes independently to node i and becomes the customer of type u with probability $p_{0(i,u)}$ $\left(\sum_{i=1}^{N}\sum_{u=1}^{M} p_{0(i,u)} = 1\right)$. Incoming signal of regime increasing and signal of regime reducing pass to node i with probabilities q_{0i}^+ and q_{0i}^- accordingly $\left(\sum_{i=1}^{N} q_{0i}^+ = 1, \sum_{i=1}^{N} q_{0i}^- = 1\right)$. After the service in node i the customer of type u passes to node j immediately with probability $p_{(i,u)(j,v)}$ as the customer of type v and with probabilities $q_{(i,u)j}^+$, $q_{(i,u)j}^-$ as the signal of increasing or reducing regime accordingly. Or it can leave the node with probability $p_{(i,u)0}$ $\left(\sum_{j=1}^{N}\sum_{v=1}^{M}(p_{(i,u)(j,v)} + q_{(i,u)j}^+ + q_{(i,u)j}^-) + p_{(i,u)0} = 1\right)$.

In each of N nodes there is the only device, which can operate in $r_i + 1$ regimes ($i = \overline{1,N}$). The state of the network is characterized by the vector $x(t) = (x_1(t), ..., x_N(t))$, where $x_i(t) = (\bar{x}_i(t), l_i(t)) = (x_{i1}(t), x_{i2}(t), ..., x_{in(i)}(t), l_i(t))$ describes the state of node i at the moment t. Here $x_{i1}(t)$ — the type of customer, which is getting service at the moment t, $x_{i2}(t)$ — the type of customer, which is the first in the queue,..., $x_{in(i)}(t)$ - the type of customer, which is the last in the queue, $n(i)$ — the number of customers in the node i, $l_i(t)$ — the regime of the node i at the moment t. States space for process $x_i(t)$ is $X_i = \{(0, l_i), (x_{i1}, l_i), (x_{i1}, x_{i2}, l_i), (x_{i1}, x_{i2}, x_{i3}, l_i), ... : x_{ik} = \overline{1,M},\ k = 1, 2, ...; l_i = \overline{0, r_i}\}$.

The time of device service of the node i has an exponential distribution with parameter $\mu_i(n(i), l_i)$. Customers are serviced in the order they arrive in the node.

We define the regime 0 as the basic regime. Switching time from some regime to another one has the exponential distribution. The node in the basic regime can pass only to the regime 1 with rate $\nu_i(\bar{x}_i, 0)$. For the states, which have the number of regime $1 \le l_i \le r_i - 1$, the node passes to regime $l_i + 1$ with rate $\nu_i(x_i)$ and to regime $l_i - 1$ with rate $\varphi_i(x_i)$. And the node passes from the regime r_i only to regime $r_i - 1$ with rate $\varphi_i(\bar{x}_i, r_i)$. While the regimes are switching in the node, the number of customers doesn't change. Switch occurs only between the neighborhood regimes.

When the signal of regime reducing incomes to the node with the regime l_i, it turns the node to the regime $l_i - 1$ and doesn't change the number of customers in the node. This signal doesn't produce any action, if the node is in the regime 0. When the signal of regime increasing incomes to the node with the regime l_i, it turns the node to the regime $l_i + 1$ and doesn't change the number of customers in the node. This signal doesn't produce any action, if the node is in the regime r_i. After changing the node regime these signals disappear.

Each regime l_i has a residence time, limited by an exponentially distributed random variable with parameter $\gamma_i(l_i)$ ($l_i = \overline{0, r_i}, i = \overline{1,N}$). After the end of the stay time in the regime l_i the device with probability $p^+(l_i)$ moves to regime $l_i + 1$, and with probability $p^-(l_i)$ moves to regime $l_i - 1$.

Then $x(t)$ is a homogeneous Markov process with states space $X = X_1 \times X_2 \times ... \times X_N$, where $X_i = \{(0, l_i), (x_{i1}, l_i), (x_{i1}, x_{i2}, l_i), ... : x_{ik} = \overline{1,M},\ k = 1, 2, ...; l_i = \overline{0, r_i}\}$.

3. ISOLATED NODE

We consider isolated node i and suppose that three independent Poisson flows come in it: the flow of customers of type u with parameter α_{iu}, the flow of signals, which increase regime of the node, with parameter β_i^+ and the flow of signals, which reduce regime of the node, with parameter β_i^-. Here α_{iu}, β_i^+, β_i^- — average rates of customers, "increasing" signals and "reducing" signals arrivals accordingly to the node i.

Traffic equations for this model are:

$$\alpha_{iu} = \lambda p_{0(i,u)} + \sum_{j=1}^{N}\sum_{v=1}^{M} \alpha_{jv} p_{(j,v)(i,u)},$$

$$\beta_i^+ = \omega^+ q_{0i}^+ + \sum_{j=1}^{N}\sum_{v=1}^{M} \alpha_{jv} q_{(j,v)i}^+,$$

$$\beta_i^- = \omega^- q_{0i}^- + \sum_{j=1}^{N}\sum_{v=1}^{M} \alpha_{jv} q_{(j,v)i}^-.$$

To reduce the calculations we introduce the following operators:
$T_u^+, T^-, S^+, S^- : X_i \to X_i$, setting

$$T_u^+(0, l_i) = (u, l_i),$$

$$T_u^+(x_i) = T_u^+(x_{i1}, \ldots, x_{in(i)}, l_i) = (u, x_{i1}, \ldots, x_{in(i)}, l_i),$$

$$T^-(x_i) = T^-(x_{i1}, l_i) = (0, l_i), \quad |x_i| = n(i) = 1,$$

$$T^-(x_i) = T^-(x_{i1}, \ldots, x_{in(i)}, l_i) = (x_{i1}, \ldots, x_{in(i)-1}, l_i), \quad |x_i| = n(i) > 1,$$

$$S^+(x_i) = S^+(x_{i1}, \ldots, x_{in(i)}, l_i) = (x_{i1}, \ldots, x_{in(i)}, l_i + 1),$$

$$S^-(x_i) = S^-(x_{i1}, \ldots, x_{in(i)}, l_i) = (x_{i1}, \ldots, x_{in(i)}, l_i - 1),$$

$T^-(x_i)$ is not defined at $x_i = (0, l_i)$, $S^+(x_i)$ — at $x_i = (\bar{x}_i, r_i)$, $S^-(x_i)$ — at $x_i = (\bar{x}_i, 0)$.
Consider also the operators describing the change of the network state:
$T_{(i,u)}^+, T_i^-, S_i^+, S_i^- : X \to X$, putting

$$T_{(i,u)}^+(x) = T_{(i,u)}^+(x_1, x_2, \ldots, x_N) = (x_1, x_2, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_N), \hat{x}_i = T_u^+(x_i);$$

$$T_i^-(x) = T_i^-(x_1, x_2, \ldots, x_N) = (x_1, x_2, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_N), \hat{x}_i = T^-(x_i);$$

$$S_i^+(x) = S_i^+(x_1, x_2, \ldots, x_N) = (x_1, x_2, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_N), \hat{x}_i = S^+(x_i);$$

$$S_i^-(x) = S_i^-(x_1, x_2, \ldots, x_N) = (x_1, x_2, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_N), \hat{x}_i = S^-(x_i).$$

Lemma. *For the reversibility of the isolated node the following conditions are necessary and sufficient*

$$[\nu_i(\bar{x}_i, l_i - 1) + \gamma_i(l_i - 1)p^+(l_i - 1) + \beta_i^+]\mu_i(n(i), l_i)[\varphi_i(T^-(\bar{x}_i, l_i)) + \gamma_i(l_i)p^-(l_i) + \beta_i^-] =$$

$$= [\nu_i(T^-(\bar{x}_i, l_i - 1)) + \gamma_i(l_i - 1)p^+(l_i - 1) + \beta_i^+]\mu_i(n(i), l_i - 1)[\varphi_i(\bar{x}_i, l_i) + \gamma_i(l_i)p^-(l_i) + \beta_i^-],$$

$$l_i = \overline{1, r_i}.$$

The proof is similar to the corresponding proof in [1].

4. STATIONARY DISTRIBUTION

Let the stationary distribution $\{p(x), x \in X\}$ of $x(t)$ exists, then the stationary state probabilities of the network satisfy the global equilibrium equations:

$$p(x)\sum_{i=1}^{N}\sum_{u=1}^{M}[\lambda p_{0(i,u)} + \mu_i(n(i), l_i)I_{(n(i)\neq 0)} +$$

$$+(\nu_i(x_i) + \gamma_i(l_i)p^+(l_i) + \omega^+ q_{0i}^+)I_{(l_i \neq r_i)} + (\varphi_i(x_i) + \gamma_i(l_i)p^-(l_i) + \omega^- q_{0i}^-)I_{(l_i \neq 0)}] =$$

$$= \sum_{i=1}^{N}[p(T_i^-(x))\lambda p_{0(i,x_{in(i)})}I_{(n(i)\neq 0)} + \sum_{u=1}^{M} p(T_{(i,u)}^+(x))\mu_i(n(i)+1, l_i)p_{(i,u)0} +$$

$$+\sum_{j=1}^{N}\sum_{v=1}^{M} p(T_{(j,v)}^+(T_i^-(x)))\mu_j(n(j)+1, l_j)p_{(j,v)(i,x_{in(i)})}I_{(n(i)\neq 0)} +$$

$$+\sum_{j=1}^{N}\sum_{v=1}^{M} p(T_{(j,v)}^+(S_i^-(x)))\mu_j(n(j)+1, l_j)q_{(j,v)i}^+ I_{(l_i \neq 0)} +$$

$$+\sum_{j=1}^{N}\sum_{v=1}^{M} p(T_{(j,v)}^+(S_i^+(x)))\mu_j(n(j)+1, l_j)q_{(j,v)i}^- I_{(l_i \neq r_i)} +$$

$$+p(S_i^-(x))[\nu_i(S^-(x_i)) + \gamma_i(l_i - 1)p^+(l_i - 1) + \omega^+ q_{0i}^+]I_{(l_i \neq 0)} +$$

$$+p(S_i^+(x))[\varphi_i(S^+(x_i)) + \gamma_i(l_i + 1)p^-(l_i + 1) + \omega^- q_{0i}^-]I_{(l_i \neq r_i)}].$$

Theorem 1. *If for all $i = \overline{1, N}$ the conditions of reversibility are true and the series converges*

$$\sum_{x \in X} q(x)\prod_{i=1}^{N}\prod_{a=1}^{n(i)}\frac{\alpha_{ix_{ia}}}{\mu_i(a, l_i)}\prod_{k=1}^{l_i}\frac{\nu_i(0, k-1) + \gamma_i(k-1)p^+(k-1) + \beta_i^+}{\varphi_i(0, k) + \gamma_i(k)p^-(k) + \beta_i^-},$$

where $(\alpha_{iu}, i = \overline{1, N}, u = \overline{1, M})$ — solution of the traffic equation,

$$q(x) = \lambda + \sum_{i=1}^{N}\mu_i(n(i), l_i)I_{(n(i)\neq 0)} + \sum_{i=1}^{N}[\nu_i(x_i) + \gamma_i(l_i)p^+(l_i) + \omega^+ q_{0i}^+]I_{(l_i \neq r_i)} +$$

$$+\sum_{i=1}^{N}[\varphi_i(x_i)+\gamma_i(l_i)p^-(l_i)+\omega^- q_{0i}^-]I_{(l_i\neq 0)},$$

then the Markov process $x(t)$ is ergodic and its stationary distribution of conditions probabilities has the product form

$$p(x)=p_1(x_1)p_2(x_2)...p_N(x_N), x\in X,$$

where $p_i(x_i)$ is determined by the relation

$$p_i(x_i)=\prod_{a=1}^{n(i)}\frac{\alpha_{ix_{ia}}}{\mu_i(a,l_i)}\prod_{k=1}^{l_i}\frac{\nu_i(0,k-1)+\gamma_i(k-1)p^+(k-1)+\beta_i^+}{\varphi_i(0,k)+\gamma_i(k)p^-(k)+\beta_i^-}p_i(0,0),$$

$$p_i(0,0)=\left[\sum_{x_i\in X_i}\prod_{a=1}^{n(i)}\frac{\alpha_{ix_{ia}}}{\mu_i(a,l_i)}\prod_{k=1}^{l_i}\frac{\nu_i(0,k-1)+\gamma_i(k-1)p^+(k-1)+\beta_i^+}{\varphi_i(0,k)+\gamma_i(k)p^-(k)+\beta_i^-}\right]^{-1}.$$

REFERENCES

1. *Малинковский, Ю. В., Нуеман А. Ю.* Мультипликативность стационарного распределения в открытых сетях с многорежимными стратегиями обслуживания // Весці НАН Беларусі. 2001. № 3. С. 129–134.
2. *Летунович Ю. Е.* Неоднородные сети с ограничением на время пребывания в режимах обслуживания // Автоматика и вычислительная техника. 2010. № 5. С. 33–41.

ESTIMATION OF LOSS PROBABILITY IN GAUSSIAN QUEUES

O. Lukashenko[1], E. Morozov[1], M. Pagano[2]

[1] *IAMR KarSC RAS,*
[2] *University of Pisa*
[1] *Petrozavodsk, Russia*
[2] *Pisa, Italy*
lukashenko-oleg@mail.ru, emorozov@karelia.ru,
m.pagano@iet.unipi.it

We discuss the application of the simulation to estimate the loss probability in a queueing system with a finite buffer, which is fed by a Gaussian input. Such estimation is based on known approximation, but less known than estimation of the overflow probability in an infinite buffer system. We focus on queues with fractional Brownian input (fBi) and Brownian input (Bi). For the Bi, regenerative simulation is applied to analyze the accuracy of the analytical approximation.

This work is supported by Russian Foundation for Basic research, project No 10-07-00017.

Keywords: fluid queues; Gaussian input; loss probability

1. INTRODUCTION

Nowadays Gaussian processes are well-recognized models to describe the traffic dynamics of a wide class of modern telecommunication networks [8]. In this work, we use an approximation of the loss probability expressed via overflow probability for infinite buffer model, which is usually more available, to develop estimation of the required loss probability. In particular, we are interested in systems with small or moderate size buffers, because the infinite buffer approximation is not a realistic assumption in real networks. In this framework, the loss rate prediction may represent a useful tool to define new call admission control algorithms in Quality of Service (Qos) supporting network architectures and for QoS routing protocols.

It is well-known that in the modern communication networks, each node is usually fed by a superposition of a large number of the individual (independent) streams. Moreover, under appropriate scaling such a superposition converges weakly to a Gaussian process or to an α-stable Levy process [6]. We limit our analysis to the Gaussian case [8]. More exactly, to motivate our interest to systems fed by fBi or Bi, consider M independent identically distributed (i.i.d) on-off sources. It then follow that, in case heavy-tailed on-off periods, appropriately scaled cumulative workload $A(Tt)$ arrived

during period $[0,\, Tt]$ converges weakly to a fractional Brownian motion (fBm) provided that [10] first, number of sources $M \to \infty$, and then scaling factor $T \to \infty$. When both on- and off-periods have finite variance, the limiting process is standard Brownian motion (Bm).

Denote by $A(s,t)$ the amount of data (input traffic) arrived into a communication node within time interval $[s,t]$, $t \geq s \geq 0$. Denote $A(t) = A(0,\, t)$ and assume that the input process $\{A(t),\, t \geq 0\}$ has stationary increments: the expected workload arrived in interval $[s,\, t]$ is $\mathsf{E}A(s,t) = m \cdot (t-s)$, where $m = \mathsf{E}A(1)$ is the mean input rate, and $VarA(s,t) = v(t-s)$ for some function v.

2. QUEUE WITH GAUSSIAN INPUT

Above given comments motivate the following definition of the Gaussian input which is used below:

$$A(t) = mt + \sqrt{am}B_H(t), \tag{1}$$

where $\{B_H(t),\, t \geq 0,\}$ is a fBm, which describes random fluctuations of the input around its linearly increasing mean. It is assumed that queueing system has single server with constant service rate C. To guarantee stability of such a system we assume that $m < C$.

The stationary overflow probability in continuous time has the following representation [5]:

$$\mathsf{P}(Q > b) = \mathsf{P}\left(\sup_{t\in[0,\infty)} (A(t) - Ct) > b\right). \tag{2}$$

This distribution is available in an explicit form in case of Bi. Namely,

$$\mathsf{P}(Q > b) = \mathsf{e}^{(-2\cdot\frac{C-m}{am}\cdot b)}. \tag{3}$$

(For more details see [9]). In the case of fBi there is no such kind of exact result, but there are some asymptotics for tail probabilities [1, 3, 7]. Above given result relates to an infinite buffer system, but our main purpose is to estimate the loss probability in a finite buffer system. Explicit or asymptotic results for finite buffer systems are much less available (at least for non Markovian systems), and simulation is often the only way to calculate the corresponding performance measures.

2.1. Loss probability in discrete time queues. Let $B_H^*(n) = B_H(n) - B_H(n-1)$, $n \geq 1$. Then queue content in infinite buffer system satisfies (in discrete time) the well-known Lindley-type recursion [8]:

$$Q(t) = \left(Q(t-1) - C + m + \sqrt{am}B_H^*(t)\right)^+,\ t = 1, 2, \ldots \tag{4}$$

where $(x)^+ = \max(0, x)$. Recursion (4) can be extended to the queue content $Q_b(t)$ in a system with buffer size $b < \infty$ as follows:

$$Q_b(t) = \min\left((Q_b(t-1) - C + m + \sqrt{m}B_H^*(t))^+, b\right),\ t = 1, 2, \ldots. \tag{5}$$

The time average loss $\mathsf{P}_\ell(b,T)$ in this system during (discrete-time) interval $[0,T]$, is naturally defined as the ratio of the amount of loss to the total amount of input during this interval, that is

$$\mathsf{P}_\ell(b,T) = \frac{\sum_{k=1}^{T}(Q_b(k-1)+m+\sqrt{am}B_H^*(k)-C-b)^+}{A(T)}. \tag{6}$$

Under stability assumption, one can expect that stationary loss ratio converges to stationary loss probability $\mathsf{P}_\ell(b)$, that is

$$\mathsf{P}_\ell(b) = \lim_{T\to\infty}\mathsf{P}_\ell(b,T) = \frac{\mathsf{E}(Q_b(n-1)+m+\sqrt{am}B_H^*(n)-C-b)^+}{m}. \tag{7}$$

(In general, the existence of such limit requires a proof because input is long-range dependent). The following expression given in [4] allows to calculate the loss probability

$$\mathsf{P}_\ell(b) \approx \frac{\mathsf{P}_\ell(a)}{\mathsf{P}(Q>a)}\mathsf{P}(Q>b), \tag{8}$$

provided there is explicit formula (or at least a satisfactory approximation) for overflow probability $\mathsf{P}(Q>x)$.

The relevance of this formula is that it allows to estimate loss probability via estimate of overflow probability in the corresponding infinite-buffer system. The latter is typically much more easy than direct calculation of the loss probability because there are explicit expressions for such a probability. Using (8), it is possible to derive estimate for the loss probability $\mathsf{P}_\ell(b)$:

$$\mathsf{P}_\ell(b) \approx \alpha e^{-(m_b/2)}, \tag{9}$$

where

$$\alpha = \frac{1}{m\sqrt{2\pi am}}\mathbf{e}^{\left(\frac{(C-m)^2}{2am}\right)\int_C^\infty (x-C)\mathbf{e}^{\left(-\frac{(x-m)^2}{2am}\right)dx},}$$

and

$$m_b := \min_{n\geq 1}\frac{(b+(C-m)n)^2}{amn+2\sum\limits_{k=1}^{n-1}am(n-k)r(k)},$$

where $r(n)$ is the autocovariance function of the increments $\{B_H^*(n)\}$, which has the following expression:

$$r(n) =: cov\,(B_H^*(0), B_H^*(n)) = 1/2[(n+1)^{2H} - 2n^{2H} + |n-1|^{2H}].$$

Note that for Bi we have exact value for m_b:

$$m_b = 4\cdot\frac{C-m}{am}\cdot b.$$

Estimate (9) is based on maximum variance asymptotic (MVA) approximation of the overflow probability $\mathsf{P}(Q>b)$ (for more details see [4]).

We note that in the case of Bi basing on explicit results (3) for continuous time queue we can obtain from (8) the following approximation for the loss probability:

$$\mathsf{P}_\ell(b) \approx \mathsf{P}_\ell(0)\mathsf{e}^{(-2\cdot\frac{C-m}{am}\cdot b)}, \quad (10)$$

where

$$\mathsf{P}_\ell(0) = \frac{1}{m\sqrt{2\pi am}} \int_C^\infty (x-C)\mathsf{e}^{\left(-\frac{(x-m)^2}{2am}\right)} dx.$$

Because Bi has independent increments, it is possible to develop regenerative analysis of the workload process $Q_b(t)$ in the system with Bi. But strictly speaking, it is impossible to give correct construction of classical regenerations for continuous time Brownian process. This difficulty is illustrated by formula (3) which implies $\mathsf{P}(Q = 0) = 0$. Hopefully, this problem can be overcome in discrete-time (slotted) setting.

For the $Bi/D/1/n$ system we can construct regeneration points for the queue content process. (More details can be found in [2].) Let $\beta_0 = 0$ and

$$\beta_{k+1} = \min\{t > \beta_k : Q(t-1) = 0,\, Q(t) > 0,\ k \geq 1\}, \quad (11)$$

where $Q(t)$ is the queue content at the end of slot t.

Now we denote by $L_b(t)$ the total load lost in time interval $[0,\, t]$. Denote by $\mathsf{E}L$ the mean load lost per regenerative cycle and let $\mathsf{E}A$ be the mean load arrived per regenerative cycle. Applying regenerative method, we obtain the following representation for the steady-state loss probability

$$P_\ell(b) = \lim_{t\to\infty} \frac{L_b(t)}{A(t)} = \frac{\mathsf{E}L}{\mathsf{E}A}.$$

3. SIMULATION RESULTS

Now we present some numerical results related to estimation of the loss probability for a single-server system $Bi/D/1/b$ with constant service rate C and finite buffer of size b.

Figure 1 compares the simulation results (based on the regenerative approach described above) for the finite buffer system $Bi/D/1/b$ with the approximations (9) and (10). The following parameters are used: $C = 1$; $m = 0.7$; $N = 10^6$ (where N denotes the number of simulated time slots), the reliability of the confidence interval is 95%. It is worth mentioning that for small buffers the analytical results are outside the confidence interval. Moreover, as expected from theory, (9) and (10) give upper bounds, i. e. conservative results for dimensioning the system.

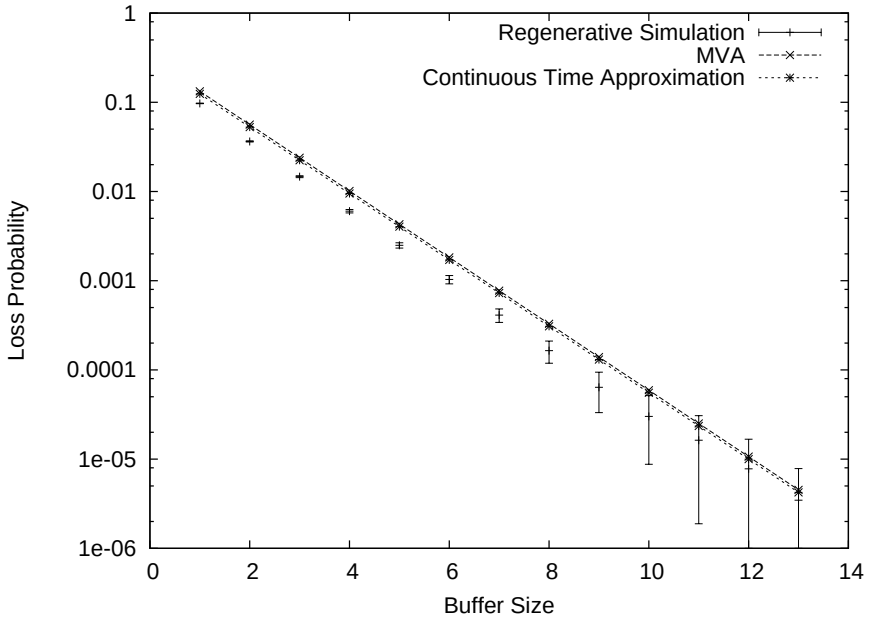

Fig. 1. Estimate of P_ℓ in Bi/D/1/b: regenerative method vs. approximation (10)

4. CONCLUSIONS

In this note, we discussed the evaluation of the loss probability in a finite buffer system fed by a Gaussian input. This has beeb done using approximation (and explicit formula) for overflow probability in the corresponding infinite buffer model. We have compared the so-called MVA approximation and the estimation using regenerative simulation in the case of Brownian input. In the following research we also plan to simulate the queue (using standard Monte Carlo algorithms) with general fBi and compare estimation of the overflow or loss probability with the approximation (9).

REFERENCES

1. *Duffield N., O'Connell N.* Large deviations and overflow probabilities for the general single server queue, with applications // Proceedings of the Cambridge Philosophical Society. 1995. V. 118. P. 363–374.
2. *Goricheva R. S., Lukashenko O. V., Morozov E. V., Pagano M.* Regenerative analysis of a finite buffer fluid queue // Proceedings of ICUMT 2010.
3. *Hüsler J., Piterbarg V. I.* Extremes of a certain class of Gaussian processes // Stochastic Processes and their Applications. 1999. V. 83. P. 257–271.
4. *Kim H. S., Shroff N. B.* Loss Probability Calculations and Asymptotic Analysis for Finite Buffer Multiplexers // IEEE/ACM Transactions on Networking, 2001. V. 9. № 6. P. 755–768.
5. *Mandjes M.* Large Deviations of Gaussian Queues // Chichester: Wiley. 2007.

6. *Mikosh T., Reshick S., Rootzen H., Stegeman A.* Is network traffic approximated by stable Levy motion or fractional Brownian motion // Annals of Applied Probability. 2002. V. 12. P. 23–68.
7. *Narayan O.* Exact asymptotic queue length distribution for fractional Brownian traffic // Advances in Performance Analysis. 1998. V. 1. P. 39–63.
8. *Norros I.* A storage model with self-similar input // Queueing Systems. 1994. V. 16. P. 387–396.
9. *Takacs L.* Combinatorial Methods in the Theory os Stochastic Processes // John Wiley&Sons. 1967.
10. *Taqqu M., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modellling // Computer Communication Review, 1997. V. 27. P. 5–23.

ИССЛЕДОВАНИЕ МАРКОВСКОЙ ДИНАМИЧЕСКОЙ RQ-СИСТЕМЫ С ВХОДЯЩИМ MMP-ПОТОКОМ ЗАЯВОК

Т. Любина*, **А. Назаров**

Томский государственный университет
Томск, Россия
* lyubina_tv@mail.ru

В качестве математической модели сети связи рассматривается однолинейная система массового обслуживания с источником повторных вызовов с входящим марковским модулированным потоком заявок, управляемая динамическим протоколом доступа. Проводится анализ данной системы и находится допредельное распределение вероятностей $p(i)$ числа заявок в источнике повторных вызовов. Обнаружено свойство стабилизации последовательности $\delta_i = p(i+1)/p(i)$, для аппроксимации распределения вероятностей, обладающего таким свойством, предложено квазигеометрическое распределение дефекта 2.

Работа выполнена в рамках Аналитической ведомственной целевой программы “Развитие научного потенциала высшей школы (2009-2010 годы)”, проект № 4761.

Ключевые слова: динамическая RQ-система, динамический протокол доступа, распределение вероятностей, квазигеометрическое распределение.

1. ПОСТАНОВКА ЗАДАЧИ

В данной работе динамической RQ-системой с входящим MMP-потоком заявок будем называть однолинейную систему массового обслуживания (СМО) с источником повторных вызовов (ИПВ) [1] с входящим марковским модулированным потоком заявок, управляемую динамическим протоколом доступа [2].

В отличие от классических моделей случайных потоков (пуассоновского и рекуррентного) математические модели специальных потоков (например, MMP, MAP, BMAP, SM) [3, 4, 5] более адекватно представляют телекоммуникационные потоки реальных данных.

Случайный поток однородных событий будем представлять в виде случайного процесса [6] $m(t)$ – числа событий потока, наступивших за время t или на интервале времени $[0,\ t)$.

Пусть задана эргодическая цепь Маркова $n(t)$, определяемая матрицей инфинитезимальных характеристик Q с элементами $q_{\nu n}$, а также набор неотрицательных чисел $\lambda_n > 0$.

Случайный поток однородных событий будем называть марковским модулированным пуассоновским потоком (ММР-потоком) [7], управляемым цепью Маркова $n(t)$, если выполнены следующие условия:

$$P\left\{m(t+\Delta t)=m+1|m(t)=m, n(t)=\nu\right\}=\lambda_\nu \Delta t+o(\Delta t),$$

$$P\left\{m(t+\Delta t)>m+1|m(t)=m, n(t)=\nu\right\}=o(\Delta t).$$

Состояниями ММР-потока будем называть состояния его управляющей цепи Маркова $n(t)$.

Таким образом, на вход системы поступает ММР-поток заявок из внешнего источника, определяемый диагональной матрицей Λ условных интенсивностей λ_n и матрицей Q инфинитезимальных характеристик $q_{\nu n}$ цепи Маркова $n(t)$, управляющей ММР-потоком. Заявка, заставшая прибор свободным, занимает его для обслуживания в течение случайного времени, распределенного по экспоненциальному закону с параметром μ. По завершении обслуживания заявка покидает прибор. Если во время обслуживания некоторой заявки поступает другая, то поступившая заявка переходит в ИПВ. Из ИПВ после случайной задержки, заявка с динамической (зависящей от состояния ИПВ) интенсивностью σ/i, вновь обращается к прибору с повторной попыткой его захвата, i – число заявок в ИПВ, то есть вероятность обращения к прибору за время Δt для любой заявки из ИПВ составляет $\frac{\sigma}{i}\Delta t+o(\Delta t)$, если в ИПВ находится i заявок. Если прибор свободен, то заявка становится на обслуживание, если же он занят, то возвращается в ИПВ. Задачей данной работы является нахождение распределения числа заявок в источнике повторных вызовов.

2. ИССЛЕДОВАНИЕ МАРКОВСКОЙ ДИНАМИЧЕСКОЙ RQ-СИСТЕМЫ С ВХОДЯЩИМ ММР-ПОТОКОМ ЗАЯВОК

Состояние системы в момент времени t определяется трёхмерной цепью Маркова $\{k(t), n(t), i(t)\}$, где $i(t)$ – число заявок в ИПВ, $n(t)$ – значения цепи Маркова, управляющей ММР-потоком, а $k(t)$ определяет состояние прибора следующим образом:

$$k=\begin{cases} 0, \text{ если прибор свободен,} \\ 1, \text{ если прибор занят обслуживанием.} \end{cases}$$

Обозначим $P\left\{k(t)=k, n(t)=n, i(t)=i\right\}=P(k,n,i,t)$ – вероятность того, что в момент времени i прибор находится в состоянии k, цепь Маркова в состоянии n и в ИПВ i заявок.

Таким образом, распределение вероятностей $P(k,n,i,t)$ удовлетворяет следующим равенствам

$$\begin{cases} P(0,n,0,t+\Delta t) = (1-\lambda_n\Delta t)(1+q_{nn}\Delta t)P(0,n,0,t) + \mu\Delta t P(1,n,0,t) + \\ + \sum\limits_{\nu\neq n} q_{\nu n}\Delta t P(0,\nu,0,t) + o(\Delta t), \\ P(1,n,0,t+\Delta t) = (1-\lambda_n\Delta t)(1-\mu\Delta t)(1+q_{nn}\Delta t)P(1,n,0,t) + \\ +\sigma\Delta t P(0,n,1,t) + \lambda_n\Delta t P(0,n,0,t) + \sum\limits_{\nu\neq n} q_{\nu n}\Delta t P(1,\nu,0,t) + o(\Delta t), \\ \ldots \\ P(0,n,i,t+\Delta t) = (1-\lambda_n\Delta t)(1-\sigma\Delta t)(1+q_{nn}\Delta t)P(0,n,i,t) + \mu\Delta t P(1,n,i,t) + \\ + \sum\limits_{\nu\neq n} q_{\nu n}\Delta t P(0,\nu,i,t) + o(\Delta t), \\ P(1,n,i,t+\Delta t) = (1-\lambda_n\Delta t)(1-\mu\Delta t)(1+q_{nn}\Delta t)P(1,n,i,t) + \sigma\Delta t P(0,n,i+1,t) + \\ +\lambda_n\Delta t P(0,n,i,t) + \lambda_n\Delta t P(1,n,i-1,t) + \sum\limits_{\nu\neq n} q_{\nu n}\Delta t P(1,\nu,i,t) + o(\Delta t), \end{cases}$$

из которых составим систему дифференциальных уравнений Колмогорова для распределения вероятностей $P(k,n,i,t)$

$$\begin{cases} \frac{\partial P(0,n,i,t)}{\partial t} = -(\lambda_n+\sigma)P(0,n,i,t) + \mu P(1,n,i,t) + \sum\limits_{\nu} P(0,\nu,i,t)q_{\nu n} \\ \frac{\partial P(1,n,i,t)}{\partial t} = -(\lambda_n+\mu)P(1,i,t) + \sigma P(0,n,i+1,t) + \lambda_n P(0,n,i,t) + \\ +\lambda_n P(0,n,i-1,t) + \sum\limits_{\nu} P(1,\nu,i,t)q_{\nu n}. \end{cases} \qquad (1)$$

Будем полагать, что система функционирует в стационарном режиме, то есть

$$P(k,n,i,t) \equiv P(k,n,i).$$

Запишем систему (1) для стационарного распределения:

$$-\lambda_n P(0,n,0) + \mu P(1,n,0) + \sum_{\nu} P(0,\nu,0)q_{\nu n} = 0, \quad i=0,$$

$$-(\lambda_n+\mu)P(1,n,0) + \lambda_n P(0,n,0) + \sigma P(0,n,1) + \sum_{\nu} P(1,\nu,0)q_{\nu n} = 0, \quad i=0,$$

$$-(\lambda_n+\sigma)P(0,n,i) + \mu P(1,n,i) + \sum_{\nu} P(0,\nu,i)q_{\nu n} = 0, \quad i\geq 1, \qquad (2)$$

$$-(\lambda_n+\mu)P(1,n,i)+\lambda_n P(0,n,i)+\lambda_n P(1,n,i-1)+\sigma P(0,n,i+1)+\sum_{\nu} P(1,\nu,i)q_{\nu n} = 0, \quad i\geq 1.$$

Для дальнейшего исследования запишем систему (2) в матричном виде. Обозначив

$$P(0,i) = \{P(0,1,i), P(0,2,i), ..., P(0,N,i)\},$$
$$P(1,i) = \{P(1,1,i), P(1,2,i), ..., P(1,N,i)\},$$

получим

$$P(0,0)(Q-\Lambda) + P(1,0)\mu = 0, \quad i=0,$$

$$P(1,0)(Q-\Lambda-\mu I)+P(0,0)\Lambda+P(0,1)\sigma=0, \quad i=0,$$
$$P(0,i)(Q-\Lambda-\sigma I)+P(1,i)\mu=0, \quad i\geq 1, \tag{3}$$
$$P(1,i)(Q-\Lambda-\mu I)+P(0,i)\Lambda+P(1,i-1)\Lambda+P(0,i+1)\sigma=0, \quad i\geq 1.$$

Чтобы решить систему (3), определим векторные характеристические функции

$$H(k,u)=\sum_{i=0}^{\infty}e^{jui}P(k,i). \tag{4}$$

Из системы (3), с учетом равенства (4), получаем следующую систему для функций $H(k,u)$:

$$\begin{cases} H(0,u)(Q-\Lambda-\sigma I)+H(1,u)\mu=-\sigma P(0,0), \\ H(0,u)(\Lambda e^{ju}+\sigma I)+H(1,u)(Q+\Lambda(e^{ju}-1)-\mu I)e^{ju}=\sigma P(0,0). \end{cases} \tag{5}$$

Из системы (5) получаем выражения для $H(0,u)$ и $H(1,u)$:

$$H(0,u)=P(0,0)\left\{\sigma I+\frac{\sigma}{\mu}e^{ju}\left(Q+\Lambda\left(e^{ju}-1\right)-\mu I\right)\right\}\left\{\sigma I(1-e^{ju})+Qe^{ju}-\right.$$
$$\left.-\frac{1}{\mu}\left(Q-\Lambda-\sigma I\right)\left(Q+\Lambda(e^{ju}-1)\right)e^{ju}\right\}^{-1}, \tag{6}$$
$$H(1,u)=-\frac{1}{\mu}\left[\sigma P(0,0)I+H(0,u)(Q-\Lambda-\sigma I)\right].$$

Перепишем уравнение (6) с учётом замен $H(k,u)=G(k,e^{ju})=G(k,x)$, $e^{ju}=x$:

$$G(0,x)=P(0,0)\left\{\sigma I+\frac{\sigma}{\mu}x\left(Q+\Lambda\left(x-1\right)-\mu I\right)\right\}\left\{\sigma I(1-x)+Qx-\right.$$
$$\left.-\frac{1}{\mu}\left(Q-\Lambda-\sigma I\right)\left(Q+\Lambda(x-1)\right)x\right\}^{-1} \tag{7}$$

и обозначим матрицы

$$A(x)=\sigma I+\frac{\sigma}{\mu}x\left(Q+\Lambda\left(x-1\right)-\mu I\right),$$

$$B(x)=\sigma I(1-x)+Qx-\frac{1}{\mu}\left(Q-\Lambda-\sigma I\right)\left(Q+\Lambda(x-1)\right)x,$$

тогда равенство (7) перепишется в следующем виде

$$G(0,x)=P(0,0)A(x)B^{-1}(x).$$

Производящая функция $G(0,x)$ определена для всех значений $x\in[0,1]$, но матрица $B(x)$ вырождена при $x=x_\nu$, где x_ν – корни уравнения

$$|B(x)|=0,$$

принадлежащие рассматриваемому интервалу $[0,\ 1]$.

Обратную матрицу $B^{-1}(x)$ запишем в виде

$$B^{-1}(x) = \frac{1}{|B(x)|} D^T(x),$$

где элементами $D(x)_{n_1 n_2}$ матрицы $D(x)$ являются алгебраические дополнения к элементам $B(x)_{n_1 n_2}$ матрицы $B(x)$.

Из равенства нулю определителя $|B(x_\nu)| = 0$ следует, что компоненты вектора $P(0,0)$ удовлетворяют однородным системам линейных алгебраических уравнений

$$P(0,0)A(x_\nu)B^T(x_\nu) = 0, \quad \forall\ x_\nu.$$

Эта система определяет значения компонент вектора $P(0,0)$ с точностью до мультипликативной постоянной, значение которой определяется условием нормировки.

3. АНАЛИЗ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ЧИСЛЕННЫМ ИНТЕГРИРОВАНИЕМ

Векторную характеристическую функцию $H(u)$ для распределения вероятностей $P(i) = P(0,i) + P(1,i)$ числа заявок в ИПВ запишем в виде

$$H(u) = H(0,u) + H(1,u) = H(0,u)\left(I - \frac{1}{\mu}(Q - \Lambda - \sigma I)\right) - \frac{\sigma}{\mu}P(0,0).$$

Тогда распределение вероятностей $p(i) = P(i)E$ числа заявок в источнике повторных вызовов определяется обратным преобразованием Фурье от скалярной характеристической функции $h(u) = Me^{jui(t)} = \sum\limits_i e^{jui}p(i) = H(u)E$:

$$p(i) = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} e^{-jui}h(u)du. \tag{8}$$

Для заданных значений параметров

$$\mu = 1, \quad \sigma = 5$$

и матриц

$$Q = \begin{pmatrix} -0,6 & 0,2 & 0,4 \\ 0,1 & -0,4 & 0,3 \\ 0,3 & 0,2 & -0,5 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 0,2 & 0 & 0 \\ 0 & 0,5 & 0 \\ 0 & 0 & 1,2 \end{pmatrix},$$

$$E = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

определяющих $h(u)$, численным интегрированием (8) получим распределение вероятностей числа заявок в ИПВ $p(i)$ (табл. 1).

Таблица 1 **Распределение вероятностей $p(i)$ числа заявок в ИПВ, $i = 0, 1, 2, \ldots$**

i	0	1	2	3	4	5	6	7	8
$p(i)$	0,2994	0,1027	0,0859	0,0728	0,0620	0,0528	0,0450	0,0384	0,0327
δ_i	0,3431	0,8358	0,8477	0,8511	0,8522	0,8526	0,8528	0,8528	0,8528
i	9	10	11	12	13	14	15	16	...
$p(i)$	0,0279	0,0238	0,0203	0,0173	0,0148	0,0126	0,0107	0,0092	...
δ_i	0,8528	0,8528	0,8528	0,8528	0,8528	0,8528	0,8528	0,8528	...

Особенность данного распределения заключается в том, что последовательность отношения $\delta_i = p(i+1)/p(i)$ достаточно быстро стабилизируются и при $i \geq 4$ принимает постоянное значение с точностью до трёх знаков после запятой.

Аналогичные результаты имеют место и для других значений параметров λ, σ и матриц Λ и Q.

Приведем следующее определение [8].

Определение. Квазигеометрическим распределением дефекта n с параметром ρ будем называть такое распределение вероятностей $p(i)$, для которого существуют ρ и C, такие что для $i \geq n$

$$p(i) = C\rho^i,$$

а при $i = n - 1$

$$p(i) \neq C\rho^i.$$

Для аппроксимации полученного в табл. 1 распределения вероятностей $p(i)$ числа заявок в ИПВ целесообразно предложить так называемое квазигеометрическое распределение $p(i)$ дефекта 2.

4. ЗАКЛЮЧЕНИЕ

Таким образом, в данной статье проведено исследование марковской динамической RQ-системы с входящим MMP-потоком заявок, то есть получено распределение вероятностей $p(i)$ числа i заявок в ИПВ. Обнаружено свойство стабилизации последовательности $\delta(i)$ для найденного распределения $p(i)$. Для аппроксимации распределения вероятностей, обладающего таким свойством, предложено квазигеометрическое распределение дефекта 2. Указанная аппроксимация является достаточно полезной при нахождении распределения вероятностей числа заявок в ИПВ приближёнными методами, такими как метод имитационного моделирования, метод асимптотического анализа, а также численными методами решения исходных систем уравнений Колмогорова для распределения вероятностей в марковских моделях.

ЛИТЕРАТУРА

1. *Назаров А. А., Терпугов А. Ф.* Теория массового обслуживания. Томск: Изд-во НТЛ, 2004.

2. *Любина Т. В., Назаров А. А.* Исследование марковской динамической RQ-системы с конфликтами заявок // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. №3 (12). С. 73–84.
3. *Cox D. R.* The analysis of non-Markovian stochastic processes // Proc. Cambr. Phil. Soc. 1955. V. 51. P. 433–441.
4. *Neuts M. F.* A versatile Markovian arrival process // Journal of Appl. Prob. 1979. V. 16. P. 764–779.
5. *Лопухова С. В.* Асимптотические и численные методы исследования специальных потоков однородных событий: дис. ... канд. физ.-мат. наук: 05.13.18; Томский гос. ун-т. Томск; 2008.
6. *Назаров А. А., Терпугов А. Ф.* Теория вероятностей и случайных процессов. Томск: Изд-во НТЛ, 2006.
7. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. 4-е изд. М.: Изд-во ЛКИ, 2007.
8. *Любина Т. В., Назаров А. А.* Аппроксимация допредельного распределения в динамической RQ-системе с конфликтами заявок // Новые информационные технологии в исследовании сложных структур: тезисы докладов Восьмой Российской конференции с международным участием. Томск: Изд-во НТЛ, 2010. С. 38.

О ПРОПУСКНОЙ СПОСОБНОСТИ ЗВЕЗДООБРАЗНОГО ФРАГМЕНТА СЕТИ

П. Михеев, С. Сущенко

Томский Государственный Университет
Томск,Россия
ssp@inf.tsu.ru

Предложена модель транзитного узла сети передачи данных, распределяющего входящий поток по нескольким исходящим каналам. Исследуется влияние качества каналов связи, различных стратегий разделения ограниченной буферной памяти транзитного узла между очередями к выходным каналам связи и распределения долей входящего трафика по исходящим направлениям на пропускную способность сетевого фрагмента.

Ключевые слова: звездообразный сетевой фрагмент, блокировки памяти, Марковская цепь, стратегии распределения буферной памяти, расщепление трафика.

1. ВВЕДЕНИЕ

В задачах анализа и проектирования компьютерных сетей важнейшим объектом исследования является звездообразный сетевой фрагмент. Всесторонний анализ звездообразного топологического образования необходим при решении задач выбора пропускных способностей, распределения потоков, реализации алгоритмов маршрутизации, разработке методов управления информационными потоками и ресурсами буферной памяти транзитных узлов передачи данных. Математические модели такой структурной конфигурации позволяют изучать пропускную способность входящих и исходящих каналов связи центрального узла коммутации с ограниченной памятью, проводить расчет емкости и оптимизацию структуры его буферного накопителя (схемы использования конечной буферной памяти для хранения очередей пакетов к выходным каналам связи) [1, 2]. Кроме того, здесь возможен анализ различных схем локального управления транзитными потоками. Одним из основных факторов, определяющих операционные характеристики сетевых структур, являются блокировки ограниченной буферной памяти узлов коммутации (на втором уровне сетевой архитектуры) и узлов маршрутизации (на третьем уровне). Основная задача исследования звездообразной конфигурации состоит в выборе наилучшей в смысле некоторого критерия (обычно вероятности блокировки или пропускной способности данного структурного образования) схемы распределения ограниченной буферной памяти центрального узла между очередями пакетов к выходным каналам связи [2]. Простейшей схемой является полное

разделение буферной памяти между каналами связи (фиксированное разбиение). Согласно этой схеме каждому выходному направлению предоставляется отдельный буферный пул. Причем, сумма размеров индивидуальных пулов равна общему объему буферной памяти узла коммутации. Противоположная данной схеме — полнодоступная стратегия, в соответствии с которой очередь к каждому из каналов связи может занимать всю имеющуюся буферную память (равнодоступная стратегия). Существует ряд промежуточных стратегий разделения памяти, наиболее общей среди которых является неполнодоступная схема с индивидуальными потолками [2]. В данной схеме каждому каналу связи выделено определенное количество индивидуальных буферов. Кроме того, имеется пул буферов общих для всех каналов, некоторую часть которых (индивидуальную для каждого канала) может занимать очередь к конкретному выходному направлению. Основным инструментом моделирования звездообразной конфигурации являются системы массового обслуживания с ограниченным накопителем, позволяющие анализировать влияние на функционирование элементов топологической структуры ближайшего сетевого окружения.

2. МОДЕЛЬ РАСЩЕПЛЕНИЯ СЕТЕВОГО ТРАФИКА

Рассмотрим звездообразный фрагмент сети, включающий $M + 1$ звено передачи данных, в котором в центральный транзитный узел по одному входящему каналу связи поступает информационный поток и распределяется по M исходящим каналам связи. Предположим, что в узле-отправителе входящего канала всегда имеются пакеты для передачи в центральный транзитный узел. Пусть обмен в каждом звене выполняется полными кадрами и организован в соответствии со стартстопным протоколом [2], согласно которому кадр считается принятым узлом-приемником, если в нем не обнаружены ошибки. При искажении информационного кадра или квитанции, подтверждающей правильность приема кадра получателем, происходит повторная передача. Предположим, что входному каналу связи выделен специальный буфер для приема кадра и анализа его на наличие ошибок. В случае корректного приема кадра, содержащийся в нем пакет переписывается в свободный буфер буферного пула выходного канала связи или (что эквивалентно) занимает данный буфер, а в качестве специального выделяется другой из того же буферного пула. При отсутствии свободных буферов в пуле выходного канала связи кадр, так же, как и при искажении, передается повторно. Такая техника гарантированного обеспечения каждого входного направления буфером для приема кадра широко используется для предупреждения «прямых» блокировок пути [2]. Полагаем, что все каналы связи имеют одинаковые физические скорости передачи данных, а узлы-отправители и узлы-получатели — одинаковое время обработки кадров при приеме и отправке. Тогда время полного цикла передачи кадра t будет одинаковым для всех звеньев рассматриваемого фрагмента. Будем считать, кроме того, что кадр, поступивший в транзитный узел в текущем цикле t, начнет передаваться по выходному каналу только в следующем цикле. Полагаем также,

что безошибочная передача кадра данных во входящем канале определяется вероятностью F, а в исходящих каналах — вероятностями F_m, $m = \overline{1, M}$. Считаем также, что весь входящий в транзитный узел поток кадров одного канала распределяется в m-й выходной канал с вероятностью B_m, $\sum_{m=1}^{M} B_m = 1$. Величины B_m определяют структуру расщепления трафика и их можно интерпретировать как доли входящего потока, направляемые в m-й выходной канал. Нетрудно видеть, что время безошибочной передачи кадра по каждому межузловому соединению является случайной величиной, кратной t. Если условия первой и повторных передач одинаковы, что, как правило, выполняется в сетях пакетной коммутации, то данная величина имеет геометрический закон распределения с параметром F во входящем канале и F_m, $m = \overline{1, M}$ — в исходящих каналах связи. Будем считать также, что для хранения пакетов в выходных очередях в транзитном узле выделен пул совместно используемой буферной памяти объема K. Размер очереди q_m к каждому выходному каналу m ограничен предельной величиной $N_m \leq K$, определяемой стратегией распределения буферной памяти между выходными каналами. Для каждого входящего пакета, направляемого в конкретный исходящий канал, выделяется буфер при условии, что выходная очередь q_m данного направления не превышает максимального размера $q_m < N_m$ и, кроме того, для очередей к выходным каналам связи выполняется ограничение $\sum_{m=1}^{M} q_m < K$, соответствующее тому, что пул свободных буферов для хранения пакетов данных не пуст. Очевидно, что в каждом конкретном случае распределения пула буферов между выходными направлениями размер очереди к m-му каналу q_m не превышает величины Q_m, удовлетворяющей условиям: $Q_m \leq N_m$ и $\sum_{m=1}^{M} Q_m = K$. В общем случае различают три стратегии распределения буферной памяти между выходными каналами связи — равнодоступную $N_m = K$, фиксированное разбиение $N_m = K/M$ (для однородных выходных каналов связи), промежуточную политику $K/M < N_m < K$, $\sum_{m=1}^{M} Q_m = K$. Основной вопрос, который приходится решать архитекторам коммуникационных систем и сетей, состоит в том как распределить совместно используемое буферное пространство для хранения очередей транзитных пакетов данных между выходными каналами связи. Поведение рассматриваемого сетевого фрагмента представимо в виде Марковской системы массового обслуживания (СМО) с дискретным временем, конечным накопителем и M обслуживающими приборами [3]. Входящий поток определяется качеством входящего канала F, а время обслуживания на каждом приборе СМО — качеством m-го исходящего канала F_m. Распределение поступающих заявок СМО по M обслуживающим приборам задается вероятностями B_m, $m = \overline{1, M}$. Динамика очередей к выходным каналам связи данной СМО в стационарных условиях описывается цепью Маркова в M-мерном пространстве. Множество возможных состояний цепи Маркова по каждому измерению определяется политикой распределения буферной памяти между исходящими каналами и не превышает величины $N_m + 1$. Для дискретной цепи Маркова с конечным числом состояний, описывающей рассматриваемую СМО в установившемся режиме, определим с учетом введенных предположений переходные вероятности π_I^J из состояния I в состояние J, где $I = i_1, i_2 \ldots i_M$; $J = j_1, j_2 \ldots j_M$;

$i_m = \overline{0, N_m}$; $j_m = \overline{0, N_m}$; $m = \overline{1, M}$ — M-разрядные номера соответственно исходного и измененного состояний с областью значений каждого разряда от 0 до N_m. Обозначим через $P_{i_1, i_2 \ldots i_M}$, $i_m = \overline{0, Q_m}$, $m = \overline{1, M}$ вероятности состояний M-мерной цепи Маркова. Важнейшей характеристикой СМО ограниченной емкости является пропускная способность. В рассматриваемом случае этот показатель интерпретируется как пропускная способность входящего звена передачи данных, нормированное значение которого определяется величиной пропущенного (обслуженного) потока:

$$Z(F, F_1 \ldots F_M, B_1 \ldots B_M) = \sum_{m=1}^{M} F_m \sum_{i_1=0}^{Q_1} \cdots \sum_{i_m=1}^{Q_m} \sum_{i_{m+1}=0}^{Q_{m+1}} \cdots \sum_{i_M=0}^{Q_M} P_{i_1, i_2 \ldots i_M}. \qquad (1)$$

3. РАВНОДОСТУПНАЯ СТРАТЕГИЯ РАЗДЕЛЕНИЯ ПАМЯТИ

Найдем функциональную зависимость вероятностей состояний СМО в стационарных условиях и индексов производительности от параметров рассматриваемого фрагмента сети. Начнем рассмотрение с простейшего случая, когда между M выходными каналами разделяется единственный буфер транзитного узла ($K = 1$). Для финальных вероятностей цепи [3] получаем следующие соотношения:

$$P_{I_m} = \begin{cases} \dfrac{\prod_{j=1}^{M} F_j (1 - F)}{\prod_{j=1}^{M} F_j (1 - F) + F \sum_{k=1}^{M} \left(B_k \prod_{j=1, j \neq k}^{M} F_j \right)}, & m = 0; \\ \dfrac{\prod_{j=1, j \neq m}^{M} F_j F B_m}{\prod_{j=1}^{M} F_j (1 - F) + F \sum_{k=1}^{M} \left(B_k \prod_{j=1, j \neq k}^{M} F_j \right)}, & m = \overline{1, M}. \end{cases}$$

Отсюда для нормированной пропускной способности (1) находим:

$$Z(F, F_1 \ldots F_M, B_1 \ldots B_M) = \sum_{m=1}^{M} F_m P_{I_m} = \frac{F \prod\limits_{m=1}^{M} F_m}{(1-F) \prod\limits_{m=1}^{M} F_m (1-F) + F \sum\limits_{m=1}^{M} \left(B_m \prod\limits_{n=1, n \neq m}^{M} F_n \right)}.$$

Из данного соотношения очевидно, что при абсолютно надежных каналах всех исходящих звеньев звездообразного сетевого фрагмента $\left(F_m = 1,\ m = \overline{1, M}\right)$ пропущенный поток определяется достоверностью передачи данных во входящем канале связи: $Z(F, 1 \ldots 1, B_1 \ldots B_M) = F$. Для статистически однородных каналов связи всего сетевого фрагмента ($F_m = F$, $m = \overline{1, M}$) выражение для пропускной способности звена передачи данных инвариантно к количеству выходных звеньев передачи данных и распределению по ним исходящего трафика:

$$Z(F, F \ldots F, B_1 \ldots B_M) = \frac{F}{2 - F}.$$

В случае, когда все исходящие каналы имеют одинаковый уровень достоверности передачи данных $\left(F_m = F_*,\ m = \overline{1, M},\ F_* \neq F\right)$ пропущенный поток инвариантен к тому, как расщепляется сетевой трафик в транзитном узле.

Продолжим рассмотрение предложенной модели сетевого фрагмента при $K = 2$, $M = 2$ и равнодоступной политике распределения буферной памяти транзитного узла ($N_m = 2$). Для статистически однородного сетевого фрагмента ($F = F_1 = F_2$) пропускная способность входящего звена передачи данных остается зависимой от структуры расщепления трафика (B_m, $m = \overline{1, 2}$). При этом равномерное расщепление трафика $B_1 = B_2 = 1/2$ обеспечивает максимальный пропущенный поток.

4. СТРАТЕГИЯ ФИКСИРОВАННОГО РАЗБИЕНИЯ ПАМЯТИ

Рассмотрим модель сетевого фрагмента с $M = 2$ выходными каналами и политике фиксированного разбиения буферной памяти равной $K = 2$ между выходными интерфейсами транзитного узла, т. е. $N_m = 1$. Переходные вероятности для этого случая приведены в таблице.

$\pi_{i_1 i_2}^{j_1 j_2}$	i_1	i_2	j_1	j_2
$1 - F$	0	0	0	0
$F_1(1 - F)$	1	0	0	0
$F_2(1 - F)$	0	1	0	0
$F_1F_2(1 - F)$	1	1	0	0
FB_1	0	0	1	0
$(1 - F_1)(1 - F + FB_1) + F_1FB_1$	1	0	1	0
F_2FB_1	0	1	1	0
$F_2(1 - F_1)(1 - F + FB_1) + F_2F_1FB_1$	1	1	1	0
FB_2	0	0	0	1
F_1FB_2	1	0	0	1
$(1 - F_2)(1 - F + FB_2) + F_2FB_2$	0	1	0	1
$F_1(1 - F_2)(1 - F + FB_2) + F_1F_2FB_2$	1	1	0	1
$(1 - F_1)FB_2$	1	0	1	1
$(1 - F_2)FB_1$	0	1	1	1
$(1 - F_1)(1 - F_2) + F_1(1 - F_2)FB_1 + F_2(1 - F_1)FB_2$	1	1	1	1

Пропускная способность при $F_1 = F_2 = F_*$ имеет вид:

$$Z(F, F_*, F_*, B_1) = FF_*\big[F_* - (1 - F)(1 - F_*)(2 - F_*) + 2\alpha(1 - F_*) - \\ -F_*(1 - F)(1 - F_*)(2 - F_*)(1 - 2FB_1 + 2FB_1^2)\big] \big/ \big[FF_* + F^3B_1(1 - B_1)(1 - F_*)^2 + \\ +F_*^2(1 - F) - \alpha F_*^2(1 - F)(1 - F_*)^2 - FF_*(1 - F)(1 - F_*)^2(1 - 2FB_1 + 2FB_1^2)\big].$$

При одинаковом качестве всех каналов связи фрагмента ($F = F_1 = F_2$) получаем следующую зависимость для пропускной способности входящего звена передачи

данных:

$$Z(F,F,F,B_1) = F\big[3F(1+2B_1-2B_1^2) - 3F^2(1+4B_1-4B_1^2) + F^3(1+8B_1-8B_1^2) - \\ -2F^4(B_1-B_1^2)\big]\big/\big[6+3B_1-3B_1^2-9F(1+B_1-B_1^2) + \\ +5F^2(1+2B_1-2B_1^2) - F^3(1+5B_1-5B_1^2) + F^4(B_1-B_1^2)\big].$$

Для детерминированного входного канала ($F=1$) имеем:

$$Z(1,F_1,F_2,B_1) = \frac{F_1F_2 + B_1(1-B_1)(F_1+F_2-2F_1F_2}{F_1(1-B_1)+F_2B_1+B_1(1-B_1)(1-F_1)(1-F_2)}.$$

Отсюда нетрудно видеть, что расщепление $B_1 = \frac{F_1(1-F_2)}{F_1(1-F_2)+F_2(1-F_1)}$ дает максимальное значение пропускной способности сетевого фрагмента. Исследование индекса пропускной способности в случае произвольных значений F и F_m свидетельствует о том, что указанное соотношение для B_1 также обеспечивает максимум операционной характеристики.

5. АНАЛИЗ СТРАТЕГИЙ РАСПРЕДЕЛЕНИЯ ПАМЯТИ

Численный анализ показывает, что для однородных исходящих каналов при равномерном распределении трафика между M исходящими направлениями ($B_m = 1/M$), $m=\overline{1,M}$ стратегия фиксированного разбиения обеспечивает бóльшие значения пропускной способности фрагмента по сравнению с равнодоступной стратегией в случае $F_m = F_* < F/M$, а при $F_m = F_* > F/M$ — имеет место обратная картина. Из сравнительного анализа следует, что ни одна из стратегий распределения буферной памяти не обеспечивает абсолютного превосходства над другими по показателю объема пропущенного потока на всей области изменения качества однородных исходящих каналов связи. Вместе с тем следует отметить, что с ростом объема буферного пула это различие нивелируется. Кроме того, промежуточная стратегия на всей области изменения параметров выходных каналов связи F_m, $m=\overline{1,M}$, либо доминирует по индексу пропускной способности над конкурирующими стратегиями, либо незначительно уступает лучшей из них. Численные исследования также показали, что в случае $K \geq M$ объем общего пропущенного потока имеет максимум по параметрам распределения трафика в исходящие каналы связи. Для малых K найдено точное аналитическое подтверждение этого факта. Экстремум наиболее ярко выражен при существенно различном качестве исходящих каналов F_m, $m=\overline{1,M}$.

ЛИТЕРАТУРА

1. *Клейнрок Л.* Вычислительные системы с очередями. М.: Мир 1979.
2. *Богуславский Л. Б.* Управление потоками данных в сетях ЭВМ. М.: Энергоатомиздат 1984.
3. *Клейнрок Л.* Теория массового обслуживания. М.: Машиностроение 1979.

ON THE INEQUALITY IN OPEN QUEUEING NETWORKS

S. Minkevičius[1,2]

[1] *Institute of Mathematics and Informatics, Akademijos 4, 08663*
[2] *Vilnius Gediminas Technical University, Sauletekio 11, 10223*
[1,2] *Vilnius, Lithuania*
stst@ktl.mii.lt

The paper is devoted to the analysis of queueing systems in the context of the network and communications theory. We investigate the inequality in an open queueing network and its applications to the theorems in heavy traffic conditions (fluid approximation, functional limit theorem, and law of the iterated logarithm) for a queue of customers in an open queueing network.

Keywords: models of information systems, queueing theory, open queueing network, heavy traffic, queue length of customers, fluid approximation, functional limit theorem, law of the iterated logarithm.

1. STATEMENT OF THE PROBLEM

The paper is devoted to the analysis of queueing systems in the context of the network and communications theory. We investigate the inequality in an open queueing network and its applications to the theorems in heavy traffic conditions (fluid approximation, functional limit theorem, and law of the iterated logarithm) for a queue of customers in an open queueing network.

Now we shall survey the papers on a queue in heavy traffic conditions. In the paper of Chen, Xinyang and Yao [3], a semi-martingale reflecting the Brownian motion approximation is developed for the performance processes such as workload, queue, and sojourn time. In the paper of Massey and Srinivasan [7], the steady-state distribution of the queue process, using tensor and Kronecker products, shows that it is of the matrix-geometric structure. J. Dai and W. Dai in [4] proved that an appropriately normalized queue process converges in distribution to a d-dimensional reflecting the Brownian motion under the heavy traffic conditions. Puhalskii [9] established moderate-deviation principles for the queue, virtual waiting time and sojourn processes. In [5], Yamada has showed that the normalized queue processes at the nodes converge in distribution to a reflected, multivariate diffusion process whose drift and diffusion coefficients are state dependent and nonsingular. In the article of Kushner and Martins [6], the authors study the problems of the pathwise average cost per unit time for controlled and uncontrolled open queueing networks in heavy traffic. In the paper of Zhang Hanqin and Xu Guang-hui [13] strong approximations for an open queueing network in heavy traffic are proved. Peterson [8] has proved that, under heavy traffic conditions, the vector pro-

cesses of total unfinished workloads converge to a multidimensional regulated Brownian motion. In the article of Reiman and Simon [11], the authors consider an open queueing network with multiple classes, priorities, "arbitrary" routing, and general service time distribution. Using a heavy traffic limit theorem for open queueing networks, Reiman [10] found the correct diffusion approximation for sojourn times in Jackson networks with a single-server station. As one can see, there are only several works designed to explore a queue in a more complicated than the classical single-server queue: tandem, multiphase queue, open queueing network (see the articles of Boxma [1, 2], Zhang Hanqin and Xu Guang-hui [13], Massey and Srinivasan [7], and Sakalauskas and Minkevičius [12]).

In this paper, we investigate an open queueing network model in heavy traffic. The service discipline is "first come, first served" (FCFS). We consider open queueing networks with the FCFS service discipline at each station and general distributions of interarrival and service times. We study the queueing network with k single server stations, each of which has an associated infinite capacity waiting room. Each station has an arrival stream from outside the network, and the arrival streams are assumed to be mutually independent renewal processes. Customers are served in the order of arrival and after service they are randomly routed to either another station in the network, or out of the network entirely. Service times and routing decisions form mutually independent sequences of independent identically distributed random variables.

The basic components of the queueing network are arrival processes, service processes, and routing processes. In particular, there are mutually independent sequences of independent identically distributed random variables $\left\{z_n^{(j)}, n \geq 1\right\}$, $\left\{S_n^{(j)}, n \geq 1\right\}$ and $\left\{\Phi_n^{(j)}, n \geq 1\right\}$ for $j = 1, 2, \ldots, k$. The random variables $z_n^{(j)}$ and $S_n^{(j)}$ are strictly positive, and $\Phi_n^{(j)}$ has support in $\{0, 1, 2, \ldots, k\}$. We define $\mu_j = \left(M\left[S_n^{(j)}\right]\right)^{-1} > 0$, $\sigma_j = D\left(S_n^{(j)}\right) > 0$, $\lambda_j = \left(M\left[z_n^{(j)}\right]\right)^{-1} > 0$, and $a_j = D\left(z_n^{(j)}\right) > 0$, $j = 1, 2, ..., k$; all of these sequences are assumed to be finite. We denote $p_{ij} = P\left(\Phi_n^{(i)} = j\right) > 0$, $i, j = 1, 2, \ldots, k$. In the context of the queueing network, the random variables $z_n^{(j)}$ function as interarrival times (from outside the network) at the station j, while $S_n^{(j)}$ is the nth service time at the station j, and $\Phi_n^{(j)}$ is a routing indicator for the nth customer served at the station j. If $\Phi_n^{(i)} = j$ (which occurs with probability p_{ij}), then the nth customer served at the station i is routed to the station j. When $\Phi_n^{(i)} = 0$, the associated customer leaves the network. To construct renewal processes generated by the interarrival and service times, we assume $z_j(0) = 0, z_j(l) = \sum\limits_{m=1}^{l} z_m^{(j)}, S_j(0) = 0, S_j(l) = \sum\limits_{m=1}^{l} S_m^{(j)}, l \geq 1, j = 1, 2, \ldots, k$. We now define $a_j(t) = \max\left(l \geq 0 : z_j(l) \leq t\right)$, $x_j(t) = \max\left(l \geq 0 : S_j(l) \leq t\right)$, $\tilde{\tau}_j(t)$ as the total number of customers routed to the jth station of the network until time t, $\tau_j(t)$ as the total number of customers after service departure from the jth station of the network until time t, $\tau_{ij}(t)$ as the total number of customers after service departure from the ith station of the network and routed to the jth station of the network until

time t, $p_{ij}^t = \frac{\tau_{ij}(t)}{\tau_i(t)}$ as a part of the total number of customers which, after service at the ith station of the network, are routed to the jth station of the network, $i, j = 1, 2, \ldots, k$ and $t > 0$. Note that this system is quite general, encompassing the tandem system, acyclic networks of $GI/G/1$ queues, networks of $GI/G/1$ queues with feedback, and an open queueing network.

First, let us denote by $Q_j(t)$ the queue of customers at the jth station of the queueing network at time t; $\beta_j = \lambda_j + \sum_{i=1}^{k} \mu_i \cdot p_{ij} - \mu_j > 0$, $\hat{\sigma}_j^2 = (\lambda_j)^3 \cdot a_j + \sum_{i=1}^{k} (\mu_i)^3 \cdot \sigma_i \cdot (p_{ij})^2 + (\mu_j)^3 \cdot \sigma_j > 0$, $j = 1, 2, \ldots, k$.

Suppose that the queue of customers in each station of the open queueing network is unlimited. All random variables are defined on the common probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

We assume that the following conditions are fulfilled:

$$\lambda_j + \sum_{i=1}^{k} \mu_i \cdot p_{ij} > \mu_j, \;\; j = 1, 2, \ldots, k. \tag{1}$$

Note that these conditions quarantee that there exists a queue of customers and it is constantly growing.

2. MAIN RESULTS

At first we will prove a lemma in which the queue length of customers in an open queueing network is estimated by a composition of renewal processes

Lemma 1. *If $Q(0) = 0$, then $|Q_j(t) - \hat{x}_j(t)| \leq \sum_{i=1}^{k} w_i(t) + \sum_{i=1}^{k} \gamma_i(t)$, where $w_j(t) = \sum_{i=1}^{k} x_i(t) \cdot |p_{ij}^t - p_{ij}|$ and $\gamma_j(t) = \sup_{0 \leq s \leq t} (x_j(s) - \tau_j(s))$, $j = 1, 2, \ldots, k$ and $t > 0$.*

Proof. By definition of the queue length of customers at the station of the network, we get that

$$\begin{aligned} Q_j(t) = \tilde{\tau}_j(t) - \tau_j(t) = \tilde{\tau}_j(t) - x_j(t) + x_j(t) - \tau_j(t) \leq \tilde{\tau}_j(t) - x_j(t) + \\ + \sup_{0 \leq s \leq t} (x_j(s) - \tau_j(s)) = \sum_{i=1}^{k} \tau_i(t) \cdot p_{ij}^t + a_j(t) - x_j(t) + \\ + \sup_{0 \leq s \leq t} (x_j(s) - \tau_j(s)) \leq \sum_{i=1}^{k} x_i(t) \cdot p_{ij}^t + a_j(t) - x_j(t) + \\ + \sup_{0 \leq s \leq t} (x_j(s) - \tau_j(s)) \leq \sum_{i=1}^{k} x_i(t) \cdot (p_{ij}^t - p_{ij} + p_{ij}) + a_j(t) - \\ - x_j(t) + \sup_{0 \leq s \leq t} (x_j(s) - \tau_j(s)) \leq \sum_{i=1}^{k} x_i(t) \cdot p_{ij} + a_j(t) - \end{aligned}$$

$$- x_j(t) + \sum_{i=1}^{k} x_i(t) \cdot |p_{ij}^t - p_{ij}| + \sup_{0 \le s \le t} (x_j(s) - \tau_j(s)) =$$

$$= \hat{x}_j(t) + \sum_{i=1}^{k} w_i(t) + \gamma_j(t) \le \hat{x}_j(t) + \left\{ \sum_{i=1}^{k} w_i(t) + \sum_{i=1}^{k} \gamma_i(t) \right\}, \quad (2)$$

where $j = 1, 2, \ldots, k$ and $t > 0$.

This implies that

$$Q_j(t) \le \hat{x}_j(t) + \left\{ \sum_{i=1}^{k} w_i(t) + \sum_{i=1}^{k} \gamma_i(t) \right\}, \quad (3)$$

$j = 1, 2, \ldots, k$ and $t > 0$ (see (2)).

On the other side, we obtain that

$$Q_j(t) \ge \tilde{\tau}_j(t) - x_j(t) = \sum_{i=1}^{k} \tau_i(t) \cdot p_{ij}^t + a_j(t) - x_j(t)$$

$$\le \sum_{i=1}^{k} (\tau_i(t) - x_i(t) + x_i(t)) \cdot p_{ij}^t + a_j(t) - x_j(t)$$

$$= \sum_{i=1}^{k} x_i(t) \cdot p_{ij}^t + a_j(t) - x_j(t) + \sum_{i=1}^{k} (\tau_i(t) - x_i(t)) \cdot p_{ij}^t$$

$$= \sum_{i=1}^{k} x_i(t) \cdot (p_{ij}^t - p_{ij} + p_{ij}) + a_j(t) - x_j(t) \quad (4)$$

$$- \sum_{i=1}^{k} (\tau_i(t) - x_i(t)) \cdot p_{ij}^t \ge \sum_{i=1}^{k} x_i(t) \cdot (p_{ij}^t - p_{ij} + p_{ij}) + a_j(t)$$

$$- x_j(t) - \sum_{i=1}^{k} x_i(t) \cdot |p_{ij} - p_{ij}^t| - \sum_{i=1}^{k} (x_i(t) - \tau_i(t))$$

$$\ge \hat{x}_j(t) - \sum_{i=1}^{k} w_i(t) - \sup_{0 \le s \le t} \left(\sum_{i=1}^{k} (x_i(s) - \tau_i(s)) \right),$$

$j = 1, 2, \ldots, k$ and $t > 0$.

This implies that

$$Q_j(t) \ge \hat{x}_j(t) - \sum_{i=1}^{k} w_i(t) - \sum_{i=1}^{k} \gamma_i(t), \quad (5)$$

$j = 1, 2, \ldots, k$ and $t > 0$.

Denote $q(t) = \sum_{i=1}^{k} w_i(t) + \sum_{i=1}^{k} \gamma_i(t),\ t > 0$. So, we obtain that

$$|Q_j(t) - \hat{x}_j(t)| \le q(t), \quad (6)$$

$j = 1, 2, \ldots, k$ and $t > 0$ (see (3) and (5)).

The proof of the lemma is completed.

□

3. APPLICATION OF THE INEQUALITY

Note that inequality (6) is the key inequality to prove several laws (fluid approximation, functional limit theorem, and law of the iterated logarithm) for a queue of customers in open queueing networks in heavy traffic conditions.

At first we present a theorem on the fluid approximation for a queue of customers in open queueing networks in heavy traffic conditions.

Theorem 1. *If conditions (1) are fulfilled, then*

$$\left(\frac{Q_1(t)}{t}; \frac{Q_2(t)}{t}; \ldots; \frac{Q_k(t)}{t}\right) \Rightarrow (\beta_1; \beta_2; \ldots; \beta_k)\,,\ 0 \le t \le 1.$$

Next, we present a theorem on the functional limit theorem for a queue of customers in open queueing networks in heavy traffic conditions.

Theorem 2. *If conditions (1) are fulfilled, then*

$$\left(\frac{Q_1(nt) - \beta_1 \cdot n \cdot t}{\hat{\sigma}_1 \cdot \sqrt{n}}; \frac{Q_2(nt) - \beta_2 \cdot n \cdot t}{\hat{\sigma}_2 \cdot \sqrt{n}}; \ldots; \frac{Q_k(nt) - \beta_k \cdot n \cdot t}{\hat{\sigma}_k \cdot \sqrt{n}}\right) \Rightarrow$$

$(z_1(t); z_2(t); \ldots; z_k(t))\,,$ *where* $z_j(t),\ j = 1, 2, \ldots, k,\ 0 \le t \le 1$ *are independent standard Winer processes.*

One of the results of the paper is the following theorem on the law of the iterated logarithm for a queue of customers in an open queueing network.

Theorem 3. *If conditions (1) are fulfilled, then*

$$P\left(\overline{\lim_{t\to\infty}} \frac{Q_j(t) - \beta_j \cdot t}{\hat{\sigma}_j \cdot a(t)} = 1\right) = 1,\ \ j = 1, 2, \ldots, k \ \ \textit{and} \ \ a(t) = \sqrt{2t \ln\ln t}.$$

Proof. The proof of these theorems is similar to that in [12], and we omit these proofs.

□

REFERENCES

1. *O. Boxma.* On a tandem queueing model with identical service times at both counters. I. // Advances in Applied Probability. 1979. 11(3). P. 616-643.
2. *O. Boxma.* On a tandem queueing model with identical service times at both counters. II. // Advances in Applied Probability. 1979. 11(3). P. 644-659.
3. *H. Chen, X. Shen and D. Yao.* Brownian approximations of multiclass open-queueing networks. // Operations Research. 2002. 50(6). P. 1032-1049.

4. *J. G. Dai and W. Dai.* A heavy traffic limit theorem for a class of open queueing networks with finite buffers. // Queueing Systems. 1999. 32(1-3). P. 5-40.
5. *K. Yamada.* Diffusion approximation for open state-dependent queueing networks in the heavy traffic situation. // Annals of Applied Probability. 1995. 5(4). P. 958-982.
6. *H. J. Kushner and L. F. Martins.* Limit theorems for pathwise average cost per unit time problems for controlled queues in heavy traffic. // Stochastics Reports. 1993. 42(1). P. 25-51.
7. *W. A. Massey and R. Srinivasan.* Steady state analysis with heavy traffic limits for semi-open networks. // Stochastic models (Ottawa, ON, 1998), CMS Conf. Proc., 26, Amer. Math. Soc., Providence, RI, 2000. P. 331-352.
8. *W. P. Peterson.* A heavy traffic limit theorem for networks of queues with multiple customer types. // Mathematics of Operations Research. 1991. 16(1). P. 90-118.
9. *A. A. Puhalskii.* Moderate deviations for queues in critical loading. // Queueing Systems. 1999. 31(3-4). P. 359-392.
10. *M. I. Reiman.* The heavy traffic diffusion approximation for sojourn times in Jackson networks. // Applied probability – computer science: the interface. Birkhauser, Boston, MA, 1982. Vol. II, 3. P. 409-423.
11. *M. I. Reiman and B. Simon.* A network of priority queues in heavy traffic: one bottleneck station. // Queueing Systems. 1990. 6(1). P. 33-58.
12. *L. L. Sakalauskas and S. Minkevičius.* On the law of the iterated logarithm in open queueing networks. // European Journal of Operational Research. 2000. 120. P. 632-640.
13. *Zhang Hanqin and Xu Guang-hui.* Strong approximations for the open queueing network in heavy traffic. // Sci. China Ser. A. 1992. 35(5). P. 521-535.

FLUID LIMIT FOR CUMULATIVE IDLE TIME IN MULTIPHASE QUEUES

S. Minkevičius[1,2], G. Kulvietis[2]

[1] *Institute of Mathematics and Informatics, Akademijos 4, 08663*
[2] *Vilnius Gediminas Technical University, Sauletekio 11, 10223*
[1], [2] *Vilnius, Lithuania*
[1] stst@ktl.mii.lt, [2] gk@fm.vtu.lt

The object of this research in the queueing theory is the Functional-Strong-Law-of-Large-Numbers (FSLLN) under the conditions of heavy traffic in Multiphase Queueing Systems (MQS). A FSLLN is known as fluid limit or fluid approximation. In this paper, the FSLLN is proved for values of important probabilistic characteristic of the MQS investigated as well as the cumulative idle time of a customer.

Keywords: heavy traffic, multiphase queues.

1. INTRODUCTION

Interest in the field of multiphase queueing systems was stimulated by the theoretical values of the results as well as by their possible applications in information and computing systems, communication networks, and automated technological processes (see, for example, [13]). The methods of investigation of single phase queueing systems are considered in [2], [3], etc. The asymptotic analysis of models of queueing systems in heavy traffic is of special interest (see, for example, [7], [8], [4], [5], etc.). The papers [9], [12] and others desribed the beginning of the investigation of diffusion approximation to queueing networks. Intermediate models - multiphase queueing systems - are considered rarer due to serious technical difficulties (see, for example, book [6]).

We present some definitions in the theory of metric spaces (see, for example, [1]). Let C be a metric space consisting of real continuous functions in [0, 1] with a uniform metric

$$\rho(x,y) = \sup_{0\le t\le 1} |x(t) - y(t)|, \; x, y \in C.$$

Let D be a space of all real-valued right-continuous functions in [0,1] having left limits and endowed with the Skorokhod topology induced by the metric d (under which D is complete and separable). Also, note that $d(x,y) \le \rho(x,y)$ for $x, y \in D$.

In this paper, we will constantly use an analog of the theorem on converging together (see, for example, [1]):

Theorem 1.

$$\begin{aligned} &\textit{Let } \varepsilon > 0 \textit{ and } \boldsymbol{X}_n,\ \boldsymbol{Y}_n,\ \boldsymbol{X} \in D. \textit{ If } \boldsymbol{P}\Big(\lim_{n\to\infty} d(\boldsymbol{X}_n, \boldsymbol{X}) > \varepsilon\Big) = 0 \\ &\textit{and } \boldsymbol{P}\Big(\lim_{n\to\infty} d(\boldsymbol{X}_n,\ \boldsymbol{Y}_n) > \varepsilon\Big) = 0, \ \textit{ then } \boldsymbol{P}\Big(\lim_{n\to\infty} d(\boldsymbol{Y}_n, \boldsymbol{X}) > \varepsilon\Big) = 0. \end{aligned} \tag{1}$$

We investigate here a k-phase multiphase queueing system (i.e., when a customer has been served at the j-th phase of the multiphase queue, he goes to the $j+1$-th phase of the multiphase queue, and after the customer has been served at the k-th phase of the multiphase queue, he leaves the multiphase queue). Let us denote by t_n the time of arrival of the n-th customer, by $S_n^{(j)}$ the service time of the n-th customer at the j-th phase of the multiphase queue, $z_n = t_{n+1} - t_n$; by $\tau_{j,n}$ the departure of the n-th customer after service at the j-th phase of the multiphase queue, $j = 1, 2, \ldots, k$.

Let interarrival times (z_n) at the multiphase queue and service times $(S_n^{(j)})$ at every phase of the multiphase queue for $j = 1, 2, \ldots, k$ be mutually independent identically distributed random variables.

Next, denote by $BI_{j,n}$ the idle time of the n-th customer at the j-th phase of the multiphase queue; $I_{j,n} = \sum\limits_{l=1}^{n} BI_{j,l}$ stands for a cummulative idle time of the n-th customer at the j-th phase of the multiphase queue, $j = 1, 2, \ldots, k$.

Suppose that the idle time of a customer at each phase of the multiphase queue is unlimited, the service principle of customers is "first come, first served" (FCFS). All random variables are defined on the common probability space (Ω, F, P).

We form such a modified multiphase queue in which $BI_{j,n} = 0,\ j = 1, 2, \ldots, k,\ n < k$. Limit distributions for the modified multiphase queue and the usual multiphase queue working in heavy traffic conditions are coincidental (see, for example, [3]). Thus, later on we will investigate only the modified multiphase queue and admit that $n \geq k$.

When $j = 1, 2, \ldots, k$, let us define

$$\delta_{j,n} = \begin{cases} S_{n-(j-1)}^{(j)} - z_n, & \text{if } n \geq k \\ 0, & \text{if } n < k. \end{cases}$$

Denote $S_{j,n} = \sum\limits_{l=1}^{n-1} \delta_{j,l}$, $S_{0,n} \equiv 0$, $\hat{S}_{j,n} = S_{j-1,n} - S_{j,n}$, $x_{j,n} = \tau_{j,n} - t_n$, $x_{0,n} \equiv 0$, $\hat{x}_{j,n+1} = x_{j,n} - \delta_{j,n+1}$, $\hat{x}_{0,n} \equiv 0$, $\alpha_j = M(z_n - S_n^{(j)})$, $\hat{y}_{j,n} = \hat{x}_{j,n} - S_{j,n}$, $j = 1, 2, \ldots, k$. Assume $S_{j,0} = 0$, $j = 1, 2, \ldots, k$.

Also assume the following condition to be fulfilled:

$$\alpha_k > \alpha_{k-1} > \cdots \alpha_1 > 0. \tag{2}$$

2. MAIN RESULT

At first we present one of the main result of paper - theorem on the FSLLN for the cumulative idle time of a customer in MQS.

Theorem 2. *If conditions (2) are fulfilled, then*

$$\left(\frac{I_{1,n}}{n};\frac{I_{2,n}}{n};\ldots;\frac{I_{k,n}}{n}\right)\Rightarrow(\alpha_1;\alpha_2;\ldots;\alpha_k).$$

Proof. At first we using that for each fixed $\varepsilon>0$ (see [11])

$$P\left(\lim_{n\to\infty}\frac{|I_{j,n}-\hat{y}_{j,n}|}{\sqrt{n}}>\varepsilon\right)=0,\ j=1,2,\ldots,k,\ n\geq k. \tag{3}$$

So,

$$P\left(\lim_{n\to\infty}\frac{|I_{j,n}-\hat{y}_{j,n}|}{n}>\varepsilon\right)=0,\ j=1,2,\ldots,k,\ n\geq k. \tag{4}$$

First we prove that

$$\frac{\hat{y}_{j,n}-\sum\limits_{i=1}^{j}\hat{S}_{i,n}}{n}=\frac{\hat{y}_{j,n}-(-S_{i,n})}{n}\Rightarrow 0,\ j=1,2,\ldots,k,\ n\geq k.$$

Using relations of [11] we obtain that

$$\begin{aligned}\hat{y}_{j,n}&=\hat{x}_{j,n}-S_{j,n}=\max_{0\leq l\leq n}(\hat{x}_{j-1,l}-S_{j,l})=\max_{0\leq l\leq n}(\hat{x}_{j-1,l}-S_{j-1,l}+S_{j-1,l}-S_{j,l})\\&=\max_{0\leq l\leq n}(\hat{y}_{j-1,l}+S_{j,l}),\ j=1,2,\ldots,k,\ n\geq k.\end{aligned} \tag{5}$$

Thus,

$$\hat{y}_{j,n}=\max_{0\leq l\leq n}(\hat{y}_{j-1,l}+S_{j,l}),\ j=1,2,\ldots,k,\ \hat{y}_{0,\cdot}\equiv 0,\quad n\geq k. \tag{6}$$

Also, we see that (see (6))

$$\begin{aligned}\hat{y}_{j,n}-\sum_{i=1}^{j}\hat{S}_{i,n}&=\max_{0\leq l\leq n}(\hat{y}_{j-1,l}+\hat{S}_{j,l})-\sum_{i=1}^{j}\hat{S}_{i,n}\geq\hat{y}_{j-1,n}+\hat{S}_{j,n}-\sum_{i=1}^{j}\hat{S}_{i,n}\\&=\hat{y}_{j-1,n}-\sum_{i=1}^{j-1}\hat{S}_{i,n}\geq\ldots\geq\hat{y}_{1,n}-\hat{S}_{1,n}=\max_{0\leq l\leq n}\hat{S}_{1,n}-\hat{S}_{1,n}\geq 0,\end{aligned} \tag{7}$$

$j=1,2,\ldots,k,\ n\geq k.$

But

$$\hat{y}_{j,n}\leq\max_{0\leq l\leq n}\hat{y}_{j-1,l}+\max_{0\leq l\leq n}\hat{S}_{j,l}=\hat{y}_{j-1,n}+\max_{0\leq j\leq l}\hat{S}_{j,l}\leq\ldots\leq\sum_{i=1}^{j}\max_{0\leq l\leq n}\hat{S}_{i,l}, \tag{8}$$

$j=1,2,\ldots,k,\ n\geq k.$

Using (7) and (8) we get that

$$0\leq\hat{y}_{j,n}-\sum_{i=1}^{j}\hat{S}_{i,n}\leq\sum_{i=1}^{j}\{\max_{0\leq l\leq n}\hat{S}_{i,l}-\hat{S}_{i,n}\}, \tag{9}$$

$j = 1, 2, \ldots, k,\ n \geq k.$

Applying (9) we achieve for each fixed $\varepsilon > 0$

$$
\begin{aligned}
P\left(\frac{|\hat{y}_{j,n} - \sum\limits_{i=1}^{j} \hat{S}_{i,n}|}{n} > \varepsilon\right) &= P\left(\frac{\hat{y}_{j,n} - \sum\limits_{i=1}^{j} \hat{S}_{i,n}}{n} > \varepsilon\right) \\
&\leq P\left(\frac{\sum\limits_{i=1}^{k}\{\max\limits_{0\leq l\leq n} \hat{S}_{i,l} - \hat{S}_{i,n}\}}{n} > \varepsilon\right) \leq P\left(\frac{\sum\limits_{i=1}^{j}\{\max\limits_{0\leq l\leq n} \hat{S}_{i,l} - \hat{S}_{i,n}\}}{n} > \varepsilon\right) \\
&\leq \sum_{i=1}^{k} P\left(\frac{\max\limits_{0\leq l\leq n} \hat{S}_{i,l} - \hat{S}_{i,n}}{n} > \varepsilon\right) = \sum_{i=1}^{k} P\left(\frac{\max\limits_{0\leq l\leq n} (-\hat{S}_{i,n-l})}{n} > \varepsilon\right), \\
&= \sum_{i=1}^{k} P\left(\frac{\max\limits_{0\leq l\leq n} (-\hat{S}_{i,l})}{n} > \varepsilon\right) \leq \sum_{i=1}^{k} P\left(\frac{\max\limits_{0\leq l\leq n} (-\hat{S}_{i,l})}{\sqrt{n}} > \varepsilon\right),
\end{aligned}
\tag{10}
$$

$j = 1, 2, \ldots, k,\ n \geq k.$

Thus, for each fixed $\varepsilon > 0$

$$
P\left(\frac{|\hat{y}_{j,n} - \sum\limits_{i=1}^{j} \hat{S}_{i,n}|}{n} > \varepsilon\right) \leq \sum_{i=1}^{k} P\left(\frac{\max\limits_{0\leq l\leq n} (-\hat{S}_{i,l})}{\sqrt{n}} > \varepsilon\right), \tag{11}
$$

$j = 1, 2, \ldots, k,\ n \geq k.$

Note (see, for example, [11]) that for each fixed $\varepsilon > 0$

$$
P\left(\lim_{n\to\infty} \frac{\max\limits_{0\leq l\leq n} (-\hat{S}_{j,l})}{\sqrt{n}} > \varepsilon\right) = 0,\ j = 1, 2, \ldots, k, \tag{12}
$$

if conditions (2) are fulfilled.

Using relation $\sum\limits_{i=1}^{j} \hat{S}_{i,n} = -S_{j,n},\ j = 1, 2, \ldots, k$, and (11)-(12) we obtain that for each fixed $\varepsilon > 0$,

$$
P\left(\lim_{n\to\infty} \frac{|\hat{y}_{j,n} - (-S_{j,n})|}{n} > \varepsilon\right) = 0,\ j = 1, 2, \ldots, k,,\ n \geq k. \tag{13}
$$

So, we get for each fixed $\varepsilon > 0$ (see (13))

$$
\begin{aligned}
&P\left(\lim_{n\to\infty} \frac{|I_{j,n} - \alpha_j \cdot n|}{n} > \varepsilon\right) \leq P\left(\lim_{n\to\infty} \frac{|I_{j,n} - \hat{y}_{j,n}|}{n} > \frac{\varepsilon}{3}\right) \\
&+ P\left(\lim_{n\to\infty} \frac{|\hat{y}_{j,n} - (-S_{j,n})|}{n} > \frac{\varepsilon}{3}\right) + P\left(\lim_{n\to\infty} \frac{|(-S_{j,n}) - \alpha_j \cdot n|}{n} > \frac{\varepsilon}{3}\right) = 0,
\end{aligned}
\tag{14}
$$

$j = 1, 2, \ldots, k,\ n \geq k.$

Thus, if conditions (2) are fulfilled, then

$$P\left(\lim_{n\to\infty} \frac{|I_{j,n} - \alpha_j \cdot n|}{n} > \varepsilon\right) = 0,\ j = 1, 2, \ldots, k,\ n \geq k. \tag{15}$$

The proof is complete. □

REFERENCES

1. *Billingsley P.* Convergence of probability measures. 1968. Wiley, New York.
2. *Borovkov A.* Stochastic processes in queueing theory. 1972. Nauka, Moscow (in Russian).
3. *Borovkov A.* Asymptotic methods in theory of queues. 1980. Nauka, Moscow (in Russian).
4. *Iglehart D.L., Whitt W.* Multiple channel queues in heavy traffic. I. // Advances in Applied Probability. 1970. V. 2. P. 150-177.
5. *Iglehart D.L., Whitt W.* Multiple channel queues in heavy traffic. II. Sequences, networks and batches. // Advances in Applied Probability. 1970. V. 2, P. 355-369.
6. *Karpelevich F.I., Kreinin A.I.* (1994). *Heavy traffic limits for multiphase queues.* American Mathematical Society, Providence.
7. *Kingman J.* On queues in heavy traffic. // J. R. Statist. Soc. 1961. V. 24, P. 383-392.
8. *Kingman J.* The single server queue in heavy traffic. // Proc. Camb. Phil. Soc. 1962. V. 57, P. 902-904.
9. *Kobyashi H.* Application of the diffusion approximation to queueing networks. // Journal of ACM. 1974. V. 21, P. 316-328.
10. *Minkevičius S.* Weak convergence in multiphase queues. // Lietuvos Matematikos Rinkinys. 1986. V. 26, P. 717-722 (in Russian).
11. *Minkevičius S.* On the full idle time in multiphase queues. // Lietuvos Matematikos Rinkinys. 2005. V. 45, P. 367-386 (in Russian).
12. *Reiman M.I.* Open queueing networks in heavy traffic. // Mathematics of Operations Research. 1984. V. 9, P. 441-459.
13. *Saati T., Kerns K.* Analytic planning. Organization of systems. 1971. Mir, Moscow (in Russian).

АНАЛИЗ СЕТЕВОЙ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ДОХОДОВ ТЕЛЕФОННОЙ КОМПАНИИ

В. Науменко[1], А. Паньков[2]

[1] *Гродненский филиал РУП "Белтелеком",*
[2] *Гродненский государственный университет*
Гродно, Республика Беларусь
[2] a.pankov@gmail.com

В работе рассматривается модель изменения доходов телефонной компании РУП "Белтелеком" от маршрутизации телефонных запросов между различными сетями мобильной связи. Моделью служит замкнутая НМ-сеть с заявками нескольких типов. Получены системы разностно-дифференциальных уравнений для нахождения ожидаемых доходов систем сети. Разработана имитационная модель рассматриваемого процесса, с помощью которой рассчитаны модельные примеры.

Ключевые слова: сеть массового обслуживания, НМ-сеть, имитационное моделирование

1. ОПИСАНИЕ МОДЕЛИ

При исследовании изменения ожидаемых доходов в сложных системах, например таких, как телефонные компании и мобильные операторы, можно использовать математические модели, построенные с использованием аппарата теории массового обслуживания (МО), в частности, модели с использованием НМ-сетей (Howard-Matalytski).

Рассмотрим модель изменения доходов Гродненского филиала РУП "Белтелеком". При наборе номера с одного сотового оператора на другой, например, с MTS на Velcom, либо наоборот, "Белтелеком" получает некоторый доход за организацию соединения двух сотовых операторов. Таким же образом происходит соединение стационарного телефона и сотового, т.е. оператора связи "Белтелеком" и одного из мобильных операторов, или наоборот. Данный процесс организован по следующему принципу. При наборе абонентом одного из мобильных операторов кода +37529 (или 8029), данный вызов обрабатывается и посылается запрос на предприятие электросвязи. На последнем срабатывает аппаратура повременного учета соединений, которая ведет посекундный учет соединений, и соответственно "Белтелеком" получает некоторый доход от одного сотового оператора, установившего соединение с другим сотовым оператором. Аналогичным образом этот процесс протекает и при организации связи между "Белтелеком" и одним из сотовых операторов либо наоборот. Схематично это можно изобразить следующим образом (рис. 1):

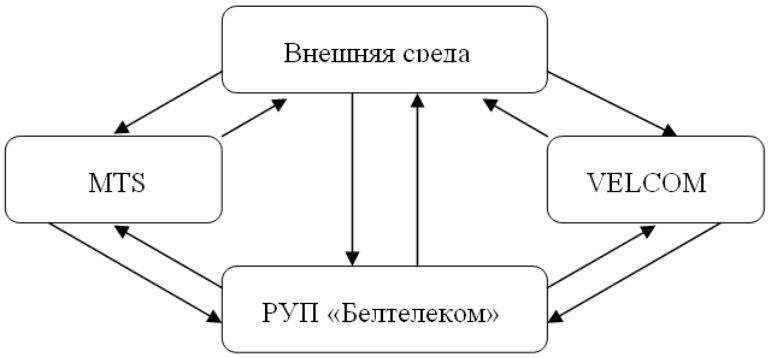

Рис. 1. Схема взаимодействия РУП "Белтелеком" с мобильными операторами

Опишем модель взаимодействия в терминах теории сетей МО [1]. Изменение доходов можно описать с помощью замкнутой НМ-сети с разнотипными заявками, состоящей из систем МО (СМО) S_1, S_2, ..., S_{n-1} (мобильные операторы), СМО S_n (РУП "Белтелеком"), а также СМО S_0 (внешняя среда). В нашем случае S_1 – это MTS, S_2 – это Velcom, S_3 – "Белтелеком", а S_0 – внешняя среда (рис. 2).

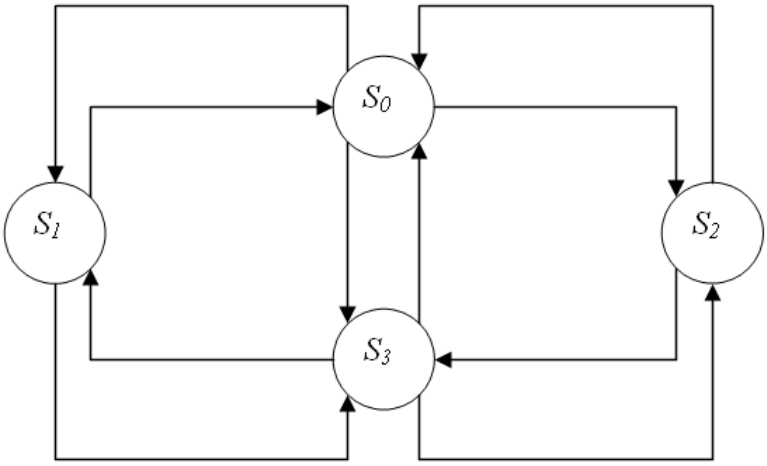

Рис. 2. Структура НМ-сети

Заявками в сети являются вызовы, поступающие от абонентов, находящихся в СМО S_0. Для нашего случая, будем предполагать, что в сети циркулирует 6 типов заявок. Если абонент желает позвонить с MTS на Velcom или наоборот, то на обслуживание поступает заявки первого и второго типа соответственно. Если абонент желает позвонить с городского телефона на MTS или наоборот, то на обслуживание поступает заявки третьего и четвертого типа соответственно. Если же абонент звонит с городского телефона на Velcom, то в сети циркулирует заявка пятого типа или наоборот – 6-го типа, когда абонент звонит с Velcom на городской телефон.

Под обслуживанием заявок будем понимать успешную/неуспешную организацию соединения операторов. Компания "Белтелеком" получаяет некоторый доход

от одного сотового оператора, установившего соединение с другим оператором, либо при организации связи между “Белтелеком” и одним из сотовых операторов или наоборот. Все заявки сети начинают и заканчивают свое обслуживание в системе S_0. Тип заявки соответствует выбору абонентом желаемого соединения.

Опишем подробно маршрут обслуживания заявок в сети, в зависимости от их типа. Если абонент выбрал тип соединения “MTS – Velcom”, то в сети МО начинает свое обслуживание заявка 1-го типа. И при переходе из системы S_1 в систему S_2 через систему S_3 – система S_3 получает некоторый доход, а система S_1 – расход за организацию соединения с другим оператором через “Белтелеком”, который установил тариф за соединение 2-ух операторов сотовой связи. Система S_1 получает лишь некоторый доход от абонента, за обеспечение соединения. Схематично переход заявки первого типа выглядит следующим образом $S_0 \to S_1 \to S_3 \to S_2 \to S_0$, т.е. заявка начинает свое обслуживание в системе S_0 и переходит в систему S_1, далее после обслуживания в системе S_1 она поступает в систему S_3 а после успешного обслуживания – в систему S_2, и далее, возвращается в систему S_0, успешно получив обслуживание в системе S_2.

Аналогично, если абонент выбрал обратное соединение типа “Velcom – MTS”, то в сети МО начинает свое обслуживание заявка 2-го типа. И при переходе из системы S_2 в систему S_1 – система S_3 получает некоторый доход. Схематично переход заявки второго типа выглядит следующим образом $S_0 \to S_2 \to S_3 \to S_1 \to S_0$.

Если же инициировано соединение типа “MTS – Белтелеком”, то в сети МО начинает свое обслуживание заявка 3-го типа. И при переходе из системы S_1 в систему S_3, система S_3 получает некоторый доход, а система S_1 – расход за организацию соединения с “Белтелеком”, который установил тариф за соединение с оператором сотовой связи. Система S_1 получает только некоторый доход от абонента, за обеспечение соединения. Схематично переход заявки третьего типа выглядит следующим образом $S_0 \to S_1 \to S_3 \to S_0$.

Если инициировано обратное соединения типа “ Белтелеком – MTS”, то в сети МО начинает свое обслуживание заявка 4-го типа. При переходе из системы S_3 в систему S_1, система S_3 получает некоторый доход, а а система S_1 – расход за организацию соединения с “Белтелеком”. Система S_1 получает лишь некоторый доход от абонента, за обеспечение соединения с опертором не сотовой связи. Схематично переход заявки первого типа выглядит следующим образом $S_0 \to S_3 \to S_1 \to S_0$. Аналогичным образом в сети функционируют заявки 5-го и 6-го типов при выборе соединения типа “Velcom – Белтелеком” ($S_0 \to S_2 \to S_3 \to S_0$) или “Белтелеком – Velcom” ($S_0 \to S_3 \to S_2 \to S_0$) соответственно.

2. СИСТЕМА РАЗНОСТНО-ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ДЛЯ НАХОЖДЕНИЯ ОЖИДАЕМЫХ ДОХОДОВ

Будем рассматривать замкнутую НМ-сеть МО $S_0, S_1, ..., S_n$, в которой обслуживаются разнотипные заявки, K_c – число заявок типа c в сети, $c = \overline{1,r}$, изображенную на рисунке 2. Системы обслуживания сети являются многолинейными, обозначим через m_i число линий обслуживания в i-й СМО,. Будем рассматривать случай, когда тип заявок не меняется при переходе из одной СМО в другую. Под состоянием сети будем понимать вектор размерности $n \times r$:

$$k = (k_{11}, k_{12}, ..., k_{1r}, k_{21}, k_{22}, ..., k_{2r}, ..., k_{n1}, k_{n2}, ..., k_{nr}),$$

где k_{ic}–число заявок типа c в системе S_i, $k_i = \sum\limits_{c=1}^{r} k_{ic}$. Обозначим через μ_{ic} – интенсивность обслуживания заявок типа c в системе S_i, $i = \overline{0,n}$, $c = \overline{1,r}$. Пусть p_{ijc} – вероятность того, что после обслуживания в i-й СМО сети заявка типа c переходит на обслуживание в j-ю СМО, $\overline{K} = (K_1, K_2, ..., K_r)$, $\quad K = \sum\limits_{c=1}^{r} K_c$, $M_{jc}(\overline{K})$ – интенсивность потока заявок типа c, выходящих из j-й СМО сети, когда в ней обслуживается K_c заявок типа c, $c = \overline{1,r}$. I_{ic} – $n \times r$-вектор из нулевых компонент, за исключением компоненты с номером $r(i-1)+c$, которая равна 1.

Через $v_n(k,t)$ обозначим полный ожидаемый доход, который получает система S_n за время t, если в начальный момент времени сеть находится в состоянии k. Предположим, что система S_n получает доход в размере $r_n(k)$ за единицу времени в течение всего периода пребывания сети в состоянии k.

Когда сеть совершает переход из состояния (k,t) в состояние $(k-I_{ic}+I_{nc}, t+\Delta t)$, она приносит системе S_n доход в размере $r(k-I_{ic}+I_{nc})$, а когда совершает переход из состояния (k,t) в состояние $(k+I_{ic}-I_{nc}, t+\Delta t)$ – доход в размере $(-R(k+I_{ic}-I_{nc}))$. В течение интервала времени Δt сеть может остаться в состоянии (k,t) либо совершить переход в состояние $(k-I_{ic}+I_{nc}, t+\Delta t)$ или $(k+I_{ic}-I_{nc}, t+\Delta t)$. Если она остается в состоянии (k,t), то доход системы S_n составит $r_n(k)\Delta t$ плюс ожидаемый доход $v_n(k,t)$, который она принесет за оставшиеся t единиц времени. В рассматриваемой модели это будет означать, что абонент выбрал иной тип соединения/разговора или иной тип предоставляемых услуг одним из операторов. Вероятность такого события равна $1-\sum\limits_{i=0}^{n} \mu_{ic} u(k_{ic})\Delta t + o(\Delta t)$, где $u(x)$ – функция Хевисайда. Если же за время Δt сеть совершает переход в состояние $(k-I_{ic}+I_{nc}, t+\Delta t)$ с вероятностью $\mu_{ic} p_{ic} u(k_i)\Delta t + o(\Delta t)$, то доход системы S_n составит $r(k-I_{ic}+I_{nc})$ плюс ожидаемый доход $v_n(k-I_{ic}+I_{nc}, t)$, который будет получен за оставшееся время, если бы начальным состоянием сети было $(k-I_{ic}+I_{nc}, t+\Delta t)$; если же за время Δt сеть совершает переход в состояние $(k+I_{ic}-I_{nc}, t+\Delta t)$ с вероятностью $\mu_{nc} p_{nic} u(k_{nc})\Delta t$, то доход составит $-R(k+I_{ic}-I_{nc})$ плюс ожидаемый доход сети за оставшееся время, если бы начальным состоянием было $(k+I_{ic}-I_{nc}, t+\Delta t)$,

$i = \overline{0, n-1}$. Тогда, используя формулу полной вероятности для математического ожидания, получаем систему разностных уравнений для дохода $v_n(k,t)$:

$$v_n(k, t+\Delta t) = \left(1 - \sum_{i=0}^{n}\sum_{c=1}^{r} m_i \mu_{ic} u(k_{ic})\Delta t\right)[r_n(k)\Delta t + v_n(k,t)] +$$

$$+\sum_{i=0}^{n-1}\sum_{c=1}^{r} m_i \mu_{ic} p_{inc} u(k_{ic})\Delta t\,[r(k - I_{ic} + I_{nc}) + v_n(k - I_{ic} + I_{nc}, t)] +$$

$$+\sum_{c=1}^{r} \mu_{nc} u(k_{nc})\Delta t \sum_{i=0}^{n-1} m_i p_{nic}\,[-R(k + I_{ic} - I_{nc}) + v_n(k + I_{ic} - I_{nc}, t)] + o(\Delta t),$$

откуда следует система разностно-дифференциальных уравнений (РДУ):

$$\frac{dv_n(k,t)}{dt} = r_n(k) - \sum_{c=1}^{r}\sum_{i=0}^{n} m_i \mu_{ic} u(k_{ic}) v_n(k,t) + \sum_{c=1}^{r}\sum_{i=0}^{n-1} m_i \mu_{ic} p_{inc} u(k_{ic}) r(k - I_{ic} + I_{nc}) -$$

$$-\sum_{c=1}^{r}\left(\mu_{nc} u(k_{nc}) \sum_{i=0}^{n-1} m_i p_{nic} R(k + I_{ic} - I_{nc}) + \sum_{i=0}^{n-1} m_i \mu_{ic} p_{inc} u(k_{ic}) v_n(k - I_{ic} + I_{nc}, t) +\right.$$

$$\left. + \mu_{nc} u(k_{nc}) \sum_{i=0}^{n-1} m_i p_{nic} v_{nc}(k + I_{ic} - I_{nc}, t)\right),$$

которую можно переписать в виде:

$$\frac{dv_n(k,t)}{dt} = r_n(k) - \sum_{c=1}^{r}\sum_{i=0}^{n} m_i \mu_{ic} u(k_{ic}) v_n(k,t) +$$

$$+\sum_{c=1}^{r}\sum_{i=0}^{n-1} m_i\,[\mu_{ic} p_{inc} u(k_{ic}) r(k - I_{ic} + I_{nc}) - \mu_{nc} u(k_{nc}) p_{nic} R(k + I_{ic} - I_{nc})] +$$

$$+\sum_{c=1}^{r}\sum_{i=0}^{n-1} m_i\,[\mu_{ic} p_{inc} u(k_{ic}) v_n(k - I_{ic} + I_{nc}, t) + \mu_{nc} u(k_{nc}) p_{nic} v_n(k + I_{ic} - I_{nc}, t)], \quad (1)$$

$$i = 0\overline{,n}, \quad c = \overline{1, r}.$$

3. РЕШЕНИЕ СИСТЕМЫ РДУ И НАХОЖДЕНИЕ ОЖИДАЕМЫХ ДОХОДОВ

Если число состояний НМ-сети невелико (<200), то систему (1) можно решить сведя ее к системе неоднородных ДУ, и решить последнюю с помощью известных методов, например методом преобразований Лапласа. Но число состояний сети на практике очень велико, так как в системе S_0 могут находиться все потенциальные абоненты области, в силу чего число уравнений в системе (1) также очень большое. Поэтому целесообразно проводить исследования методом имитационного моделирования (ИМ).

ИМ представляет собой способ проведения на ЭВМ вычислительных экспериментов с математическими моделями, имитирующими поведение реальных объектов, процессов и систем в течение заданного периода. Одной из наиболее привлекательных особенностей метода ИМ также является возможность прогнозировать и планировать поведение системы в будущем.

Имитационная модель для описанной сети была разработана с помощью библиотеки SymPy. Пакет SimPy – это открытая программная библиотека для имитационного моделирования различных процессов с дискретными событиями, написанная на языке программирования Python. С помощью этой модели можно найти ожидаемые доходы систем НМ-сети (рис. 2), а также построить графики изменения доходов на интервале моделирования. Использование такой модели, а также ПО, позволяет оценить ожидаемые доходы Гродненского филиала РУП “Белтелеком” и осуществлять их прогнозирование на различных временных периодах.

4. МОДЕЛЬНЫЙ ПРИМЕР

Рассмотрим замкнутую сеть МО с доходами, состоящую из 4 СМО. S_0, S_1, S_3, S_4. Пусть в сети циркулируют 100 заявок первого типа, 22 заявки второго типа, 34 заявки третьего типа, 10 заявок четвертого типа, 20 заявок 5-го типа и 40 заявок 6-го типа. Всего в сети циркулирует $K = 226$ заявок. Пусть времена обслуживания заявок в системах распределены по экспоненциальному закону с параметрами $\mu_0 = 0.67$, $\mu_1 = 0.11$, $\mu_2 = 0.83$ и $\mu_3 = 0.6$. Будем рассматривать случай, когда доходы от переходов сети между состояниями, являются тарифами, установленными за организацию соединения двух сотовых операторов. Пусть также в начальный момент времени денежные средства во всех системах МО сети отсутствуют. Число линий обслуживания в СМО: $m_1 = 10$, $m_2 = 10$, $m_3 = 40$. Матрицы вероятностей переходов заявок 1-го, 2-го, 3-го, 4-го, 5-го и 6-го типов в системах сети имеют

следующий вид соответственно: $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0,33 & 0,33 & 0,33 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$ и $\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0,33 & 0,33 & 0,33 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

При помощи имитационного моделирования построены графики для ожидаемых доходов систем, которые приведены на рисунках 3 – 6.

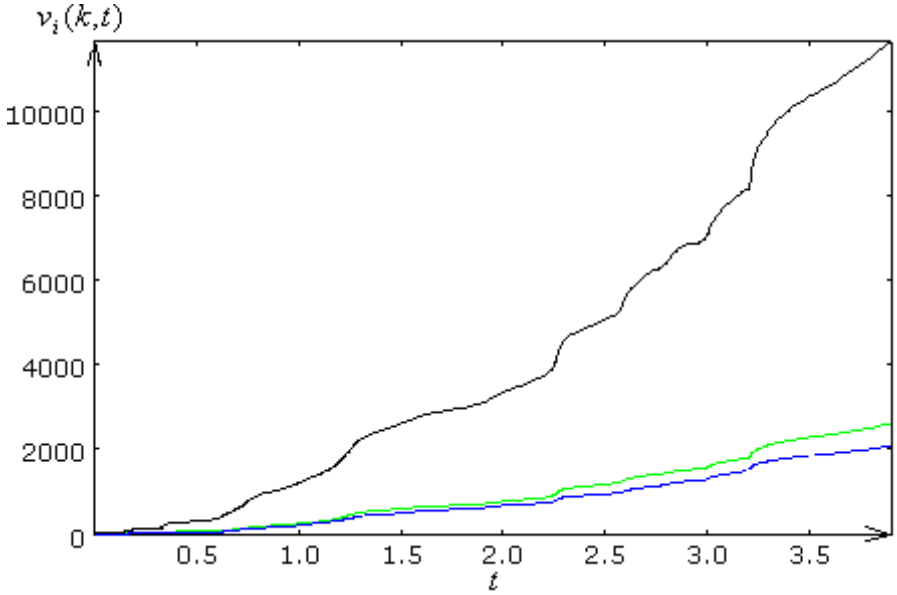

Рис. 3. Изменение доходов систем S_1, S_2, S_3 в зависимости от времени t, $\overline{i=1,3}$

ЛИТЕРАТУРА

1. *Паньков А.* Анализ сетей массового обслуживания с центральной системой в переходном режиме // Массовое обслуживание: потоки, системы, сети: материалы междунар. науч. конф. BWWQT–2007. г. Гродно. 29 янв. – 1 февр., 2007. Вып. 19. С. 201–206.

ИССЛЕДОВАНИЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С НЕОГРАНИЧЕННЫМ ЧИСЛОМ ОБСЛУЖИВАЮЩИХ ПРИБОРОВ И ВХОДЯЩИМ РЕКУРРЕНТНЫМ ПОТОКОМ

А. Назаров, И. Семенова*

Томский государственный университет
Томск, Россия
*inna_ac@mail.ru

В данной работе проведено исследование системы массового обслуживания с неограниченным числом обслуживающих приборов и входящим рекуррентным потоком. Записаны моменты первого, второго и третьего порядков числа занятых приборов в блоке обслуживания. Проведено исследование данной системы методом асимптотического анализа до третьего порядка.

Работа выполнена в рамках аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы (2009-2010 годы)" проект № 4761.

Ключевые слова: рекуррентный поток заявок, метод моментов, асимптотический анализ.

1. ВВЕДЕНИЕ

Исследованию рекуррентного потока событий посвящено большое количество работ. В книге Д.Кокса, В. Смита [2] подробно исследуется рекуррентный поток событий как основная и наиболее простая модель теории восстановления. Данная теория получила многочисленные применения в ряде направлений исследования – теории надежности, теории массового обслуживания, теории запасов и многих других приложениях. Результаты теории восстановления являются мощным средством исследования как теоретических, так и прикладных проблем. Часто применение того или иного факта теории восстановления позволяет очень просто получать результаты, которые сложно и трудно получать при другом подходе.

2. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим систему массового обслуживания с неограниченным числом приборов. На вход системы поступает рекуррентный поток заявок, заданный функцией распределения $A(x)$ – длин интервалов между моментами наступления событий рассматриваемого потока. Продолжительности обслуживания различных заявок

стохастически независимы, одинаково распределены и имеют экспоненциальное распределение с параметром μ. Поступившая заявка занимает любой из свободных приборов, завершив обслуживание, заявка покидает систему.

Для рассматриваемой системы обозначим $i(t)$ – число занятых приборов. В силу того, что процесс $i(t)$ является немарковским, определим процесс $z(t)$ как длину интервала от момента t до момента наступления следующего события в рекуррентном потоке, тогда случайный двумерный процесс $\{z(t), i(t)\}$ является марковским. Найдем распределение вероятностей этого процесса

$$P\left(z,i,t\right)=P\{z\left(t\right)<z,i\left(t\right)=i\}.$$

Для распределения вероятностей $P(z,i,t)$ составим систему уравнений Колмогорова в стационарном режиме

$$\frac{\partial P\left(z,i\right)}{\partial z}-\frac{\partial P\left(0,i\right)}{\partial z}-i\mu P\left(z,i\right)+\left(i+1\right)P\left(z,i+1\right)+\frac{\partial P\left(0,i-1\right)}{\partial z}A\left(z\right)=0. \quad (1)$$

Применяя (1), составим систему уравнений, определяющих характеристические функции $H\left(z,u\right)=\sum\limits_{i=0}^{\infty}e^{jui}P\left(z,i\right)$, для $H(z,u)$ получим уравнение

$$\frac{\partial H\left(z,u\right)}{\partial z}+\frac{\partial H\left(0,u\right)}{\partial z}\left(e^{ju}A\left(z\right)-1\right)-\mu j\left(e^{-ju}-1\right)\frac{\partial H\left(z,u\right)}{\partial u}=0, \quad (2)$$

решение $H(z,u)$ которого удовлетворяет условию

$$H\left(z,0\right)=\sum_{i=0}^{\infty}P\left(z,i\right)=R\left(z\right),$$

где $R(z)$ – стационарное распределение вероятностей значений случайного процесса $z(t)$. Известно, что распределение $R(z)$ имеет вид

$$R\left(z\right)=\lambda\int\limits_{0}^{z}\left(1-A\left(x\right)\right)dx, \quad (3)$$

где $\lambda=\frac{1}{a}$, $a=\int\limits_{0}^{\infty}xdA\left(x\right)$, в частности имеет место равенство

$$\left.\frac{\partial R\left(z\right)}{\partial z}\right|_{z=0}=\lambda.$$

3. МЕТОД МОМЕНТОВ

3.1. Момент первого порядка. Из свойств характеристических функций имеем [1] $\left.\frac{\partial H(z,u)}{\partial u}\right|_{u=0}=jm_1\left(z\right)$. Продифференцируем уравнение (2) по u, получим

$$\frac{\partial^2 H\left(z,u\right)}{\partial z\partial u}+\frac{\partial^2 H\left(0,u\right)}{\partial z\partial u}\left(e^{ju}A\left(z\right)-1\right)+\frac{\partial H\left(0,u\right)}{\partial z}je^{ju}A\left(z\right)-$$

$$-\mu j\left(-j\right)e^{-ju}\frac{\partial H\left(z,u\right)}{\partial u}-\mu j\left(e^{-ju}-1\right)\frac{\partial^2 H\left(z,u\right)}{\partial u^2}=0, \tag{4}$$

откуда при $u=0$ имеем

$$\frac{\partial m_1\left(z\right)}{\partial z}+\frac{\partial m_1\left(0\right)}{\partial z}\left(A\left(z\right)-1\right)+\lambda A\left(z\right)-\mu m_1\left(z\right)=0. \tag{5}$$

Дальнейшее решение будем проводить с помощью преобразования Лапласа-Стильтьеса

$$\varphi_1\left(\alpha\right)=\int\limits_0^\infty e^{-\alpha z}dm_1\left(z\right),\quad A^*\left(\alpha\right)=\int\limits_0^\infty e^{-\alpha z}dA\left(z\right),$$

выполнив в (5) преобразования Лапласа-Стильтьеса, получим равенство

$$\left(\mu-\alpha\right)\varphi_1\left(\alpha\right)=\frac{\partial m_1\left(0\right)}{\partial z}\left(A^*\left(\alpha\right)-1\right)+\lambda A^*\left(\alpha\right), \tag{6}$$

откуда

$$\varphi_1\left(\alpha\right)=\frac{1}{\mu-\alpha}\left[\frac{\partial m_1\left(0\right)}{\partial z}\left(A^*\left(\alpha\right)-1\right)+\lambda A^*\left(\alpha\right)\right]. \tag{7}$$

Обозначим в (6) $\alpha{=}\mu$, получим равенство

$$\frac{\partial m_1\left(0\right)}{\partial z}=\lambda\frac{A^*\left(\mu\right)}{1-A^*\left(\mu\right)}. \tag{8}$$

В силу того, что $\varphi_1\left(0\right)=\int\limits_0^\infty dm_1\left(z\right)=m_1\left(\infty\right)-m_1\left(0\right)=m_1\left(\infty\right)$, из равенства (7), и учитывая (8), запишем

$$m_1\left(\infty\right)=\varphi_1\left(0\right)=\frac{1}{\mu}\left[\lambda\frac{A^*\left(\mu\right)}{1-A^*\left(\mu\right)}\cdot\left(A^*\left(0\right)-1\right)+\lambda A^*\left(0\right)\right].$$

Используя то, что $A^*\left(0\right)=\int\limits_0^\infty dA\left(z\right)=A\left(\infty\right)-A\left(0\right)=1$, получим

$$m_1\left(\infty\right)=\frac{\lambda}{\mu}.$$

3.2. Момент второго порядка. Аналогично моменту первого порядка запишем равенство, определяющее момент третьего порядка числа занятых приборов

$$m_2\left(\infty\right)=\frac{\lambda}{\mu}\left\{\frac{1}{1-\mathrm{A}^*\left(\mu\right)}\right\}.$$

3.3. Момент третьего порядка. Момент третьего порядка числа занятых приборов имеет вид

$$m_3(\infty) = \frac{\lambda}{\mu}\left\{\frac{1}{1-\mathrm{A}^*(\mu)}\left[1+2\frac{\mathrm{A}^*(2\mu)}{1-\mathrm{A}^*(2\mu)}\right]\right\}.$$

Для более детального исследования многолинейной системы массового обслуживания с входящим рекуррентным потоком воспользуемся основным уравнением для характеристических функций (2), которое будем решать методом аимптотического анализа в условии растущего времени обслуживания, то есть при $\mu \to 0$, что позволит найти асимптотическое распределение вероятностей числа занятых приборов в рассматриваемой системе.

4. МЕТОД АСИМПТОТИЧЕСКОГО АНАЛИЗА

4.1. Асимптотика первого порядка. Для нахождении асимптотики первого порядка обозначим $\mu = \varepsilon$, и в уравнении (2) выполним замены $u = \varepsilon w$, $H(z,u) = F_1(z,w,\varepsilon)$, для $F_1(z,w,\varepsilon)$ получим уравнение

$$\frac{\partial F_1(z,w,\varepsilon)}{\partial z} + \frac{\partial F_1(0,w,\varepsilon)}{\partial z}\left(e^{j\varepsilon w}A(z)-1\right) - j\left(e^{-j\varepsilon w}-1\right)\frac{\partial F_1(z,w,\varepsilon)}{\partial w} = 0. \qquad (9)$$

Теорема 1. *Предельное, при $\varepsilon \to 0$, значение $F_1(z,w)$ решения $F_1(z,w,\varepsilon)$ уравнения (9) имеет вид*

$$F_1(z,w) = R(z)\cdot \mathbf{e}^{\{jw\lambda\}},$$

где вектор функция $R(z)$ и параметр λ определены выше.

Доказательство. В уравнении (9) выполним предельный переход при $\varepsilon \to 0$, получим уравнение

$$\frac{\partial F_1(z,w)}{\partial z} + \frac{\partial F_1(0,w)}{\partial z}\left(A(z)-1\right) = 0,$$

решение $F_1(z,w)$ которого запишем в виде

$$F_1(z,w) = R(z)\,\Phi_1(w), \qquad (10)$$

где $R(z)$ определено выше, а функцию $\Phi_1(w)$ определим следующим образом. В уравнении (9) выполним предельный переход, при $z \to \infty$, затем подставив сюда произведение (10) получим равенство определяющее функцию $\Phi_1(w)$

$$\Phi_1(w) = \mathbf{e}^{\{jw\lambda\}}.$$

□

В силу произведения (10) и обратных замен для характеристической функции величины $i(t)$ в стационарном режиме, запишем

$$Me^{jui(t)} = H(\infty,u) = \mathbf{e}^{\left\{ju\frac{\lambda}{\mu}\right\}}.$$

Полученное равенство будем называть асимптотикой первого порядка характеристической функции числа занятых приборов системы с входящим рекуррентным потоком [3].

4.2. Асимптотика второго порядка. Для нахождении асимптотики второго порядка в уравнении (2) выполним замену $H(z,u) = H_2(z,u)\mathbf{e}^{\left\{ju\frac{\lambda}{\mu}\right\}}$, обозначим $\mu = \varepsilon^2$, $u = \varepsilon w$, $\quad H_2(z,u) = F_2(z,w,\varepsilon)$. Тогда для F_2(z,w,ε) получим уравнение

$$\frac{\partial F_2(z,w,\varepsilon)}{\partial z} + \frac{\partial F_2(0,w,\varepsilon)}{\partial z}\left(e^{j\varepsilon w}A(z) - 1\right) + j\varepsilon\left(1 - e^{-j\varepsilon w}\right)\frac{\partial F_2(z,w,\varepsilon)}{\partial w} -$$
$$- \lambda\left(1 - e^{-j\varepsilon w}\right)F_2(z,w,\varepsilon) = 0. \tag{11}$$

Теорема 2. *Предельное, при $\varepsilon \to 0$, значение $F_2(z,w)$ решения $F_2(z,w,\varepsilon)$ уравнения (11) имеет вид*

$$F_2(z,w) = R(z)\cdot\mathbf{e}^{\left\{\frac{(jw)^2}{2}\kappa_2\right\}},$$

где величина κ_2 определяется равенством

$$\kappa_2 = \lambda + \frac{\partial f_2(0)}{\partial z}, \tag{12}$$

а вектор функция $f_2(z)$ удовлетворяет условию $f_2(\infty)E = 0$ и является решением уравнения

$$\frac{\partial f_2(z)}{\partial z} + \frac{\partial f_2(0)}{\partial z}(A(z) - 1) + \frac{\partial R(0)}{\partial z}A(z) - \lambda R(z) = 0. \tag{13}$$

Доказательство. Решение $F_2(z,w,\varepsilon)$ уравнения (11) запишем в виде разложения

$$F_2(z,w,\varepsilon) = \Phi_2(w)\{R(z) + j\varepsilon w f_2(z)\} + O(\varepsilon^2). \tag{14}$$

Здесь вектор функция $R(z)$ определена выше, вектор функция $f_2(z)$ удовлетворяет условию $f_2(\infty)E = 0$ и будет определена ниже, так же как и функция $\Phi_2(w)$.

Подставляя (14) в равенство (11) получим

$$\frac{\partial R(z)}{\partial z} + j\varepsilon w\frac{\partial f_2(z)}{\partial z} + \left\{\frac{\partial R(0)}{\partial z} + j\varepsilon w\frac{\partial f_2(0)}{\partial z}\right\}(A(z) - 1) +$$
$$+ j\varepsilon w\frac{\partial R(0)}{\partial z}A(z) - \lambda j\varepsilon w R(z) = O(\varepsilon^2). \tag{15}$$

В силу того, что выполняется равенство $\frac{\partial R(z)}{\partial z} + \frac{\partial R(0)}{\partial z}(A(z) - 1) = 0$, равенство (15) перепишем в виде

$$\frac{\partial f_2(z)}{\partial z} + \frac{\partial f_2(0)}{\partial z}(A(z) - 1) - \lambda R(z) + \frac{\partial R(0)}{\partial z}A(z) = 0,$$

которое совпадает с (13) и определяет функцию $f_2(z)$.

Для того, чтобы найти скалярную функцию $\Phi_2(w)$ в уравнении (11) устремим $z \to \infty$, разложим в ряд экспоненты и выполним замену (14), получим равенство

$$\varepsilon^2 w\frac{\partial\Phi_2(w)}{\partial w} = 2\lambda\frac{(j\varepsilon w)^2}{2}\Phi_2(w) + (j\varepsilon w)^2\Phi_2(w)\frac{\partial f_2(0)}{\partial z} + \mathrm{O}\left(\varepsilon^3\right) \quad ,$$

в котором выполним предельный переход при $\varepsilon \to 0$, получим равенство

$$\Phi_2(w) = \mathbf{e}^{\left\{\frac{(jw)^2}{2}\kappa_2\right\}},$$

где $\kappa_2 = \lambda + \frac{\partial f_2(0)}{\partial z}$ и совпадает с (12) $\square$

Для характеристической функции величины $i(t)$ в стационарном режиме, запишем

$$Me^{jui(t)} = H_2(\infty, u) = \mathbf{e}^{\left\{ju\frac{\lambda}{\mu} + \frac{(ju)^2}{2}\frac{\kappa_2}{\mu}\right\}}.$$

Полученное равенство будем называть асимптотикой второго порядка характеристической функции числа занятых приборов.

4.3. Асимптотика третьего порядка. Аналогичным образом для асимптотики третьего порядка была сформулированиа и доказана следующая теорема.

Теорема 3. *Предельное, при $\varepsilon \to 0$, значение $F_3(z, w)$ решения $F_3(z, w, \varepsilon)$ уравнения*

$$\frac{\partial F_3(z, w, \varepsilon)}{\partial z} + \frac{\partial F_3(0, w, \varepsilon)}{\partial z}\left(e^{j\varepsilon w}A(z) - 1\right) + j\varepsilon^2\left(1 - e^{-j\varepsilon w}\right)\frac{\partial F_3(z, w, \varepsilon)}{\partial w} -$$

$$-\left(1 - e^{-j\varepsilon w}\right)(\lambda + j\varepsilon w\kappa_2)F_3(z, w, \varepsilon) = 0$$

имеет вид

$$F_3(z, w) = R(z) \cdot \mathbf{e}^{\left\{\frac{(jw)^3}{6}\kappa_3\right\}},$$

где величина κ_3 определяется равенством

$$\kappa_3 = \kappa_2 + \frac{\partial f_2(0)}{\partial z} + \frac{\partial f_3(0)}{\partial z}$$

а вектор функция $f_3(z)$ является решением уравнения

$$\frac{\partial f_3(z)}{\partial z} + \frac{\partial f_3(0)}{\partial z}(A(z) - 1) + 2\frac{\partial f_2(z)}{\partial z}A(z) - 2\kappa_2 R(z) = 0.$$

Для характеристической функции величины $i(t)$ в стационарном режиме, запишем

$$Me^{jui(t)} = H_3(\infty, u) = \mathbf{e}^{\left\{ju\frac{\lambda}{\mu} + \frac{(ju)^2}{2}\frac{\kappa_2}{\mu} + \frac{(ju)^3}{6}\frac{\kappa_3}{\mu}\right\}}.$$

Полученное равенство будем называть асимптотикой третьего порядка характеристической функции чмсла занятых приборов.

5. ЗАКЛЮЧЕНИЕ

Таким образом, найдены моменты первого, второго и третьего порядков, а также проведено исследование системы методом асимптотического анализа до третьего порядка.

Численная реализация позволит определить условие применимости асимптотического метода [4] для анализа системы массового обслуживания с входящим рекуррентным потоком и неограниченным числом обслуживающих приборов.

ЛИТЕРАТУРА

1. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. М.: КомКнига, 2005. 228 с.
2. *Кокс Д., Смит В.* Теория восстановления. М.: Сов. радио, 1967. 298 с.
3. *Назаров А. А., Моисеева С. П.* Метод асимптотического анализа в теории массового обслуживания Томск: Изд-во НТЛ. 2006. 112 с.
4. *Назаров А. А., Семенова И. А.* Исследование RQ-систем методом асимптотических семиинвариантов // Вестник Томского государственного университета. Серия управление, вычислительная техника и информатика 2010. № 3 (12). с. 85-96.

ИССЛЕДОВАНИЕ НЕМАРКОВСКОЙ RQ-СИСТЕМЫ С КОНФЛИКТОМ ЗАЯВОК

А. Назаров, Е. Судыко*

Томский государственный университет
Томск, Россия
* ESudyko@yandex.ru

В работе рассмотрена немарковская RQ-система (Retrial queue) с конфликтами заявок, при возникновении которых обе заявки, обслуживаемая и поступившая, переходят в источник повторных вызовов. Предложен метод асимптотических семиинвариантов для исследования RQ-системы.

Ключевые слова: RQ-система, источник повторных вызовов, конфликты заявок.

1. ВВЕДЕНИЕ

В данной статье рассмотрим однолинейную немарковскую RQ-систему (Retrial queues) с конфликтами заявок , на вход которой поступает простейший поток заявок с интенсивностью λ. Требование, заставшее прибор свободным, занимает его для обслуживания в течение случайного времени, имеющего произвольную функцию распределения $B(x)$. Если прибор занят, то поступившая и обслуживаемая заявки вступают в конфликт и обе переходят в источник повторных вызовов (ИПВ), где осуществляют случайную задержку, продолжительность которой имеет экспоненциальное распределение с параметром σ. Из ИПВ после случайной задержки заявка вновь обращается к прибору с повторной попыткой его захвата. Если прибор свободен, то заявка из ИПВ занимает его для обслуживания.

2. ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

Пусть $i(t)$- число заявок в ИПВ, $z(t)$ - длина интервала от момента t до момента окончания обслуживания заявки, стоящей на приборе в момент времени t, а $k(t)$ - определяет состояние прибора следующим образом:
$k(t) = 0$, если прибор свободен;
$k(t) = 1$, если прибор находится в состоянии обслуживания заявки.

Компонента $z(t)$ определяется только в те моменты, когда $k(t) = 1$. Обозначим

$$P\{k(t) = 0, i(t) = i\} = P(0, i, t)$$

вероятность того, что прибор в момент времени t свободен ($k(t) = 0$)и в источнике повторных вызовов находится i заявок, а

$$P\{k(t) = 1, z(t) < z, i(t) = i\} = P(1, z, i, t)$$

вероятность того, что прибор в момент времени t занят ($k(t) = 1$), до конца обслуживания заявки, стоящей на приборе в момент времени t, осталось время, меньшее, чем z и в источнике повторных вызовов находится i заявок.

Процесс $\{k(t), z(t), i(t)\}$ изменения во времени состояний описанной системы является марковским.

Для распределения вероятностей $P(k, z, i, t)$ состояний $\{k, z, i, t\}$ рассматриваемой RQ-системы составим, систему дифференциальных уравнений Колмогорова, которую запишем для стационарного случая

$$(\lambda + i\sigma)P(0, i) = \lambda P(1, \infty, i-2) + (i-1)\sigma P(1, \infty, i-1) + \frac{\partial P(1, 0, i)}{\partial z},$$
$$(\lambda + i\sigma)P(1, z, i) = \lambda P(0, i)B(z) + (i+1)\sigma P(0, i+1)B(z) + \frac{\partial P(1, z, i)}{\partial z} - \frac{\partial P(1, 0, i)}{\partial z}. \tag{1}$$

Систему (**??**) для функции $H(0, u) = \sum\limits_i \mathbf{e}^{jui} P(0, i)$, $H(1, z, u) = \sum\limits_i \mathbf{e}^{jui} P(1, z, i)$ можно записать следующим образом

$$\lambda H(0, u) - j\sigma \frac{\partial H(0, u)}{\partial u} = \lambda \mathbf{e}^{2ju} H(1, u) - j\mathbf{e}^{ju}\sigma \frac{\partial H(1, u)}{\partial u} + \frac{\partial H(1, 0, u)}{\partial z},$$
$$\lambda H(1, z, u) - j\sigma \frac{\partial H(1, z, u)}{\partial u} = \lambda H(0, u)B(z) -$$
$$-j\mathbf{e}^{-ju}\sigma \frac{\partial H(0, u)}{\partial u} B(z) + \frac{\partial H(1, z, u)}{\partial z} - \frac{\partial H(1, 0, u)}{\partial z}. \tag{2}$$

Эта система будет называться основной для дальнейшего исследования методом асимптотических семиинвариантов.

3. АСИМПТОТИЧЕСКИЙ АНАЛИЗ

3.1. Асимптотика первого порядка. Для нахождения асимптотики первого порядка в уравнении (2) выполним следующие замены

$$\sigma = \varepsilon, \quad u = \varepsilon w, \quad H(0, u) = F_1(0, w, \varepsilon), \quad H(1, z, u) = F_1(1, z, w, \varepsilon). \tag{3}$$

Тогда уравнения системы (2) примут вид

$$\lambda F_1(0, w, \varepsilon) - j\frac{\partial F_1(0, w, \varepsilon)}{\partial w} = \lambda \mathbf{e}^{2j\varepsilon w} F_1(1, w, \varepsilon) - j\mathbf{e}^{j\varepsilon w} \frac{\partial F_1(1, w, \varepsilon)}{\partial w} + \frac{\partial F_1(1, 0, w, \varepsilon)}{\partial z},$$
$$\lambda F_1(1, z, w, \varepsilon) - j\frac{\partial F_1(1, z, w, \varepsilon)}{\partial w} = \lambda F_1(0, w, \varepsilon)B(z) -$$
$$- j\mathbf{e}^{-j\varepsilon w} \frac{\partial F_1(0, w, \varepsilon)}{\partial w} B(z) + \frac{\partial F_1(1, z, w, \varepsilon)}{\partial z} - \frac{\partial F_1(1, 0, w, \varepsilon)}{\partial z}. \tag{4}$$

Теорема 1. *Предельное при* $\varepsilon \to 0$ *значение вектор-строки* $\{F_1(0,w), F_1(1,z,w)\}$ *решения* $\{F_1(0,w,\varepsilon), F_1(1,z,w,\varepsilon)\}$ *системы* (4) *имеет вид*

$$F_1(0,w) = R(0)\Phi_1(w) = R(0)\mathbf{e}^{jw\kappa_1},$$

$$F_1(1,z,w) = R(1,z)\Phi_1(w) = R(1,z)\mathbf{e}^{jw\kappa_1}, \tag{5}$$

где $R(0)$ *и* $R(1,z)$ *определяются равенствами*

$$R(0) = \frac{1}{2 - B^*(\lambda+\kappa_1)}, \tag{6}$$

$$R(1,z) = \Big(1 - \mathbf{e}^{(\lambda+\kappa_1)z}\Big)\frac{B^*(\lambda+\kappa_1)}{2 - B^*(\lambda+\kappa_1)} - \frac{\mathbf{e}^{(\lambda+\kappa_1)z}}{2 - B^*(\lambda+\kappa_1)}\int_0^z B(s)d\mathbf{e}^{-(\lambda+\kappa_1)s}, \tag{7}$$

а величина κ_1 *является решением нелинейного скалярного уравнения*

$$B^*(\lambda+\kappa_1)(2\lambda+\kappa_1) = 2\lambda, \tag{8}$$

где

$$\int_0^\infty \mathbf{e}^{-sx}dB(x) = B^*(s).$$

3.2. Асимптотика второго порядка. Для нахождения асимптотики второго порядка в уравнениях основной системы (2) выполним следующие замены

$$H(0,u) = H_2(0,u)\mathbf{e}^{\frac{ju}{\sigma}\kappa_1}, \qquad H(1,z,u) = H_2(1,z,u)\mathbf{e}^{\frac{ju}{\sigma}\kappa_1},$$

$$\sigma = \varepsilon^2, \quad u = \varepsilon w, \quad H_2(0,u) = F_2(0,w,\varepsilon), \quad H_2(1,z,u) = F_2(1,z,w,\varepsilon), \tag{9}$$

для неизвестных функций $\{F_2(0,w,\varepsilon), F_2(1,z,w,\varepsilon)\}$ получим уравнения

$$(\lambda+\kappa_1)F_2(0,w,\varepsilon) - j\varepsilon\frac{\partial F_2(0,w,\varepsilon)}{\partial w} = \mathbf{e}^{j\varepsilon w}F_2(1,w,\varepsilon)\Big(\mathbf{e}^{j\varepsilon w}\lambda + \kappa_1\Big) -$$
$$-j\mathbf{e}^{j\varepsilon w}\varepsilon\frac{\partial F_2(1,w,\varepsilon)}{\partial w} + \frac{\partial F_2(1,0,w,\varepsilon)}{\partial z},$$
$$(\lambda+\kappa_1)F_2(1,z,w,\varepsilon) - j\varepsilon\frac{\partial F_2(1,z,w,\varepsilon)}{\partial w} = \Big(\lambda + \mathbf{e}^{-j\varepsilon w}\kappa_1\Big)F_2(0,w,\varepsilon)B(z) -$$
$$-j\mathbf{e}^{-j\varepsilon w}\varepsilon\frac{\partial F_2(0,w,\varepsilon)}{\partial w}B(z) + \frac{\partial F_2(1,z,w,\varepsilon)}{\partial z} - \frac{\partial F_2(1,0,w,\varepsilon)}{\partial z}. \tag{10}$$

Теорема 2. *Предельное при* $\varepsilon \to 0$ *значение вектор-строки* $\{F_2(0,w), F_2(1,z,w)\}$ *решения* $\{F_2(0,w,\varepsilon), F_2(1,z,w,\varepsilon)\}$ *системы* (10) *имеет вид*

$$F_2(0,w) = R(0)\Phi_2(w) = R(0)\mathbf{e}^{\frac{(jw)^2}{2}\kappa_2},$$

$$F_2(1,z,w) = R(1,z)\Phi_2(w) = R(1,z)\mathbf{e}^{\frac{(jw)^2}{2}\kappa_2}, \tag{11}$$

где величина κ_2 определяется равенством

$$\kappa_2 = \frac{R(0)\kappa_1\Big(\theta + (\lambda+\theta)B^*(\theta)\Big) + R(1)\Big(\lambda\theta + (\lambda+\theta)^2\Big)}{R(0)\Big(\theta - (\lambda+\theta)(B^*(\theta)-1)\Big) - R(1)\Big(\lambda + 2\theta\Big) - R^*(\theta)(\lambda+\theta)}, \tag{12}$$

а $\theta = \lambda + \kappa_1$ и $R^*(\theta)$ имеет вид:

$$R^*(s) = -s^2 \int_0^\infty \mathbf{e}^{-sx} x\Big\{R(0)(1-B(x)) - R(1)\Big\}dx. \tag{13}$$

Функция распределения числа заявок в источнике повторных вызовов имеет вид

$$h_2(u) = \mathbf{e}^{ju\frac{\kappa_1}{\sigma} + \frac{(ju)^2}{2}\frac{\kappa_2}{\sigma}}.$$

4. ЧИСЛЕННЫЕ РАСЧЕТЫ

При заданной функции распределения $B(x)$ времени обслуживания заявки на приборе решая уравнение (8), найдем значение κ_1 . Из равенства (12) определим значение скалярной величины κ_2.

4.1. Экспоненциальное время обслуживания заявки на приборе. В случае экспоненциального времени обслуживания заявки на приборе преобразование Лапласа-Стилтьеса $B^*(\lambda+\kappa_1)$ от функции $B(x)$ будет выглядеть следующим образом:

$$B^*(\lambda+\kappa_1) = \int_0^\infty \mathbf{e}^{-(\lambda+\kappa_1)s} dB(s) = \int_0^\infty \mathbf{e}^{-(\lambda+\kappa_1)s}\mu\mathbf{e}^{-\mu s}ds = \frac{\mu}{\mu+\lambda+\kappa_1}. \tag{14}$$

Определим значения параметров в виде

$$\lambda = 0,4; \quad j = \sqrt{-1}; \quad \mu = 1. \tag{15}$$

Получим следующие значения асимптотических семиинвариантов при помощи выражений (8) и (12) совместно с (13), $\kappa_1 = 1,6$ и $\kappa_2 = 8,8$ значения которых совпадают со значениями асимптотических семиинвариантов, полученных в [2].

4.2. Детерминированное время обслуживания заявки на приборе. В случае детерминированного времени обслуживания заявки на приборе преобразование Лапласа-Стилтьеса $B^*(\lambda+\kappa_1)$ от функции $B(x)$ будет выглядеть следующим образом:

$$B^*(\lambda+\kappa_1) = \int_0^\infty \mathbf{e}^{-(\lambda+\kappa_1)s} dB(s) = \mathbf{e}^{-(\lambda+\kappa_1)} \tag{16}$$

Подставляя это равенство в уравнение (8), получаем, что в зависимости от выбранного значения параметра λ имеет место три ситуации: когда уравнение (8) для определения скалярной величины κ_1 имеет 2 корня, 1 корень и не имеет ни одного корня. Эти ситуации можно наглядно представить при помощи таблицы значений параметра λ следующим образом:

λ	Число корней уравнения (8)
$\lambda < 0.2119$	2 корня
$\lambda = 0.2119$	1 корень
$\lambda > 0.2119$	нет корней

Если уравнение (8) имеет 2 корня, то большему из них соответствуют отрицательные значения κ_2 из равенств (12), в силу чего можно сделать вывод об отсутствии стационарного режима в системе. Об отсутствии стационарного режима свидетельствует и ситуация, когда уравнение (8) не имеет ни одного корня.

Таким образом стационарный режим в системе достигается только при наличии одного корня в уравнении (8), однако следует отметить, что такая ситуация не представляет практического интереса, так как реализуется лишь при единственном значении $\lambda = 0.2119$.

5. ЗАКЛЮЧЕНИЕ

Таким образом, в работе проведено исследование немарковской RQ-системы с конфликтами заявок, при возникновении которых обе заявки, обслуживаемая и поступившая, переходят в источник повторных вызовов, методом асимптотического анализа в условии растущего времени задержки заявки в источнике повторных вызовов. Сформулированы теоремы получения уравнения для определения семиинварианта первого порядка и формулу для получения семиинварианта второго порядка. Сформулировано следствие, определяющее значение характеристической функции числа заявок в источнике повторных вызовов.

ЛИТЕРАТУРА

1. *Назаров А. А., Терпугов А. Ф.* Теория массового обслуживания. Томск: Изд-во НТЛ, 2004.
2. *Назаров А. А., Судыко Е. А.* Метод асимптотических семиинвариантов для исследования математической модели сети случайного доступа // Проблемы передачи информации. 2010. № 1. С. 94–111.

SOME MARTINGALE RELATIONS FOR M/G/1 RETRIAL QUEUE

H. Oukid[1], A. Aissani[2]

[1] *University of Blida,*
[2] *USTHB*
[1] *Blida, Algeria)*
[2] *USTHB, Algeria)*
aaissani@usthb.dz

Abstract

In this paper we apply the martingale method due to Baccelli and Makowski for analysing the $M/G/1$ retrial queue. Using the recursive equation of the process embedded at depart epochs, we construct a discrete-time martingale stopped at the first passage time where the system becomes empty. We derive the stability condition and study the busy period of this system.

Keywords:Retrial queues, Embedded Markov Chain, Martingale, Busy Period.

1. INTRODUCTION

During the past years, an important research effort has been devoted to retrial queues due to their specificity and their ability in modeling several systems. Our queuing system is characterized by the phenomenon that an arriving customer who finds the server busy upon arrival is obliged to leave the service area and repeat his demand after some time called retrial time. Between trials, a blocked customer that remains in a retrial group is said to be orbit. The major analytic results and techniques used in retrial systems area are summarized in Yang and Templeton [2] and Falin [5] survey papers. We also refer to the synthesis presented by A.Aissani [4]. Regarding analysis or optimization for which stability problems are studied, often under specific assumptions, the martingale method represents an alternative approach; although, this approach was not very often used Queuing Theory. The purpose of this work is to enlarge the scope of applicability of martingale method to M/G/1 retrial systems.

2. THE MATHEMATICAL MODEL

We consider an $M/G/1$ retrial queuing system where the primary customers arrive according to a Poisson process with arrival rate λ and the service times are independent and identically distributed with arbitrary probability distribution $B(.)$ and Laplace - Stieljes transforms $B^*(.), Re(s) \geq 0$. The time between successive repeated attempts are exponentially distributed with rate ν. If the server is free at the instant of a primary call, the arriving customer begins service immediately and leaves the system

after service completion. Otherwise, if the customer finds the server busy, then it enters orbit and becomes a source of repeated calls (secondary call source). We define the number of customers in the system at time t to be $X(t)$. Note that this process is not a Markov chain since it depends on the history of the process and not just upon the current state. So, we consider the discrete time process $\{X_n, n \geq 1\}$, where X_n is the number of customers in the system as seen by the nth departing customer. More precisely,

$$\{X_n = X(t_n); n = 1, 2, ...\}$$

where t_n is the departure time of the nth customer. According to the results established in [6],

$$X_{n+1} = X_n + A_{n+1} - \delta_{X_n}, n \geq 1. \tag{1}$$

where A_n is the number of primary calls during the nth customer service time; $\delta_{X_n} = 1$ If the $(n+1)th$ customer comes from the orbit and $\delta_{X_n} = 0$ otherwise. The random variables A_n are mutually independent and their joint distribution is given by

$$P(A_n = k) = \int_0^\infty \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t), k = 0, 1, 2, ...; n \geq 1. \tag{2}$$

The generating function of this distribution is

$$a(z) = \sum_{k=0}^{\infty} z^k P(A_n = k) = B^*(\lambda - \lambda z)), 0 \leq z \leq 1. \tag{3}$$

where $B*(s)$ is the Laplace-Stieltjes transform of the service time probability distribution. Let $(\Omega, \Im, P)$ be the basic probability space, where $\Im$ is the σ-algebra generated by the input parametric sequences. We also define the increasing sequence of sigma-algebras $\Im_n$ by $\Im_n = \sigma\{A_m : 0 < m < n\}$ generated by the sequence of events $\{A_m\}$. Consequently the random variables X_n are $\Im_n$-measurable and the random variable $A_{n+1}$is independent from the F_n σ-algebras. With the above notations and using the properties of the conditional expectations

$$E(z^{X_{n+1}}/\Im_n) = z^{X_n - \delta_{X_n}} a(z) a.s. \tag{4}$$

3. THE MARTINGALE

We can define a martingale $M_n(z)$ with filtration $(\Im_n)$ by

$$M_0(z) = z^{X_0}$$

$$M_n(z) = z^{X_n} \frac{z \sum_{k=0}^{n-1} \delta_{X_k}}{a(z)^n}, 0 \leq z \leq 1 \tag{5}$$

is an integrally positive martingale.

Proof. It is not difficult to see that the sequence $\{M_n(z), n \in N\}$ is a positive martingale since

$$E(M_{n+1}(z)/\Im_n) = M_n(z) \tag{6}$$

Moreover, from the Martingale theorem [7]it is integrable.

The quantity $\rho = \lambda\alpha$ is called the traffic intensity i.e. the mean number of arrivals per mean service time. □

4. STABILITY OF THE M/G/1 RETRIAL SYSTEM

We first study the instability of the M/G/1 retrial system.

Theorem 1. *Under the assumption $\rho > 1$, the $M/G/1$ retrial system is unstable, and additionally, we have a.s.*

$$lim_{n\to\infty} X_n = \infty$$

Proof. For every $0 < z \leq 1$ and every $n \in N$, the relation (5) implies that

$$E(Z^{X_{n+1}}/\Im_n) \leq \frac{a(z)}{z} z^{X_n} a.s. \tag{7}$$

Under the assumption $\rho > 1$ and since $a(.)$ is convex, then according to Takacs lemma [6] (p .47), we can find z_0 such $c_0 = \frac{a(z_0}{z_0} < 1$. Consequently $E(z_0^{X_n}/\Im_n) \leq C_0 z_0^{X_n} \leq z_0^{X_n}$ a.s., which proves that the sequence $\{z_0^{X_n}, n \in N\}$ is an a.s. majorized by a constant $a = 1$ positive super martingale. According to [6] (th II-2-9, p26), this sequence converges then a.s. On the other hand, by using the Dominated Convergence Theorem, we obtain

$$lim_n E(z_0^{X_n}) = E(lim_n z_0^{X_n} \tag{8}$$

By recurrence on n, we deduce that

$$E(z_0^{X_n}) \leq c_0^n E(z_0^{X_0} \leq c_0^n \tag{9}$$

Passing to the limit when n tends to the infinity, we get

$$lim_n E(z_0^{X_n}) = E(lim_n z_0^{X_n}) = 0 \tag{10}$$

which implies that $lim_n E(z_0^{X_n}) = 0$ a.s. for $0 < z_0 < 1$. This leads to the result. □

The case where $\rho \leq 1$ remains for study.

We consider σ as an arbitrary stopping time for $\Im_n$ and we define the random variable $\nu(\sigma)$ as the first instant after the time $\sigma$where the system comes back to its empty state. That is

$$\nu(\sigma) = inf\{n \geq 1 : X_{\sigma+n} = 0\}, if \sigma < \infty, \tag{11}$$

$$\nu(\sigma) = 0, otherwise,$$

with the convention that $inf\{\emptyset\} = +\infty$. The following theorem formulates a conservation law for the recall $M/G/1$ system and that is essential to prove the stability result.

Theorem 2. *For $0 < z \leq 1$ and under the assumption $\rho \leq 1$,*

$$E(1_{[\sigma<\infty,\nu(\sigma)<\infty]}\frac{z\sum_{k=\sigma}^{\tau(\sigma)-1}\delta_{X_k}}{a(z)\nu(\sigma)}/\Im_\sigma) = 1_{[\sigma<\infty]}z^{X_\sigma}a.s. \tag{12}$$

Proof. We consider $\tau(\sigma) = \sigma + \nu(\sigma)$ a stopping time for $\{\Im_t, t \in N\}$. For every $t \geq 0$, $\tau(\sigma)\wedge t$ and $\sigma\wedge t$ will still be the stopping times. It is clear that $\forall\sigma\wedge t \leq \tau(\sigma)\wedge t$. Since $\{M_n(z)\}_0^\infty$ is an integrable positive martingale, then according to [7] (cor.IV-2-6,p.67), we have for every $0 < z \leq 1$ and $t \in N$

$$E(M_{\tau(\sigma)\wedge t}(z)/\Im_{\sigma\wedge t}) = M_{\sigma\wedge t}(z)a.s. \tag{13}$$

By using proposition II-1-3 of [7] (p21) and the $\Im_{\sigma\wedge t}$ measurability of the event $[\sigma < t]$, the equality (13) writes up as follow

$$E(1_{[\sigma<t]}M_{\tau(\sigma)\wedge t}(z)/\Im_{\sigma\wedge t}M_{\sigma\wedge t}(z)a.s. \tag{14}$$

On the other hand, since $a(.)$ is convex and $\rho \leq 1$ then according to Takacs lemma [6](p47], for $0 < z \leq 1$ we have $z < a(z)$. Consequently, for every $0 < z \leq 1$ and $t \in N$,

$$0 \leq 1_{[\sigma<t]}z_{X_{\tau(\sigma)\wedge t}}\frac{z^{\sum_{k=0}^{\tau(\sigma)\wedge t}}}{a(z)^{\tau(\sigma)\wedge t}} \leq 1_{[\sigma<t]}z^{-\sum_{k=0}^{\tau(\sigma)\wedge t}-1}\delta_{X_k} \tag{15}$$

□

5. MAIN RESULT

We formulate, now the main result,

Theorem 3. *Assume that $\rho = -\lambda B^{*'}(0) \leq 1$, and the sequence of service times forms a renewal sequence, then there exists a sequence $\{\tau_n(n)\}_1^\infty$ of a.s. finite $\Im_n$-stopping times such that $X_{\tau_n} = 0$ on $\{\tau_n < \infty\}$ and $\tau_n + 1 \leq \tau_{n+1}, \forall n \in N*$. Additionally, if $\nu_{n+1} := \tau_{n+1} - \tau_n foraln > 0$, then the random variables $\{\tau_n(n)\}_2^\infty$ form an i.i.d. sequence independent of τ_1 and*

$$E(\nu_{n+2}) = \frac{\Psi(1)}{1-\rho}, if\rho < 1, \tag{16}$$

$$E(\nu_{n+2}) = \infty, if\rho = 1, \tag{17}$$

where $\Psi(1) = \mathbf{e}$
$(\lambda\frac{}{\nu\int_0^1\frac{1-a(u)}{a(u)-u}du})$.

The proof of the theorem 3 follows the methodology of Baccelli and Makowski [1] using the above Martingale relations and the fact that for the M/G/1 Retrial Queue, the busy period satisfies the equation

$$E(\nu(\sigma)) = I_\infty exp(\frac{\lambda}{\nu}\int_0^1\frac{1-a(y)}{a(y)-y}dy$$

REFERENCES

1. *Baccelli F., Makowski A. M.* Direct martingale arguments for stability: The M/G/1 case // systems and Control Letters. 1994. № 6. P. 181–186.
2. *Yang T., Templeton J. G.* A survey of retrial queues // Queuing Syst., № 2. 1987. P. 201–233.
3. *Falin G. I., Templeton J. G.* Retrial Queues // Chapman and Hill, N.Y. 1997.
4. *Aissani A.* A survey of Retrial Queues // Actes des Journйes de Statistiques Appliquйes, 1994. P. 1–11.
5. *Falin G. I.* A Survey of Retrial Queues // Queueing Syst. 1990. V. 7. P. 127–168.
6. *Takacs L.* Introduction to the Theory of Queues // Oxford University Press, London 1962.
7. *Neveu J.* Martingales a temps discret // Masson, Paris. 1972.

ANALYSIS AND OPTIMIZATION OF EXPONENTIAL QUEUEING NETWORKS BY MEANS OF PARALLEL CALCULATIONS

A. Pankov, E. Koluzaeva

Grodno State University
Grodno, Belarus
koluzaeva@gmail.com

The method of a finding of mean characteristics of an open exponential queueing network (QN) and solution of the problem of coasts optimization in such network by means of parallel calculations is considered. The software for solution of the formulated problems, developed with usage of technology of programming MPI and libraries PETSc is described.

*Keywords:*queueing network, parallel calculations.

1. INTRODUCTION

Now one of the most complicated problems in the world is considered the solution of tasks which need big calculation resources. During research of the time-probability characteristics of the QN with usage of classical methods [1] it is often necessary to solve systems of great number of difference-differential equations which they satisfy. The number of equations in these systems corresponds to number of the network states and grows exponentially with adding in the network at least one message or new queueing system (QS) (for closed networks). For open QN the situation especially worse cause the number of states of such networks is countable. Therefore evaluation of characteristics of QN by regular computer machines can vary from several hours (for numerical methods) to several days and more (for analytical methods). The next step in development of QN research methods is usage of high performance computing (HPC) for calculations. However, for parallel computing one needs to develop special software. That is time and effort consuming process.

The problem of finding of the exponential QN characteristics as well as solution of the network cost optimization problem using parallel computations are considered in the paper.

2. FINDING OF MEAN CHARACTERISTICS OF JACKSON OPEN NETWORK

Let's consider an open QN consisting of n QS S_1, S_2, ..., S_n. For unification we enter QS S_0, corresponding to an external environment. Each system has FIFO service

discipline. An arrival flow of messages in the network is Poisson with rate λ. Service times of messages in various QS are independent and don't depend on flows arriving in the systems; thus the service time of the messages in i-th QS has an exponential distribution with parameter $\mu_i(l)$, where l – number of messages in QS. The message serviced by i-th QS, passes into j-th QS with probability p_{ij} , and leaves the network with probability p_{i0} , $\sum\limits_{j=0}^{n} p_{ij} = 1$, $i = \overline{1,n}$.

From the Jackson expansion theorem [2] follows that final stationary probabilities of Jacson's network states look like

$$P(k) = \prod_{i=1}^{n} P_i(k_i), \tag{1}$$

where $P_i(k_i)$, $P_i(0)$ are defined by following:

$$P_i(k_i) = \frac{\lambda e_i}{\mu_i(k_i)} P_i(k_i - 1) = \frac{\lambda e_i}{\mu_i(k_i)} \frac{\lambda e_i}{\mu_i(k_i - 1)} P_i(k_i - 2) = ... = \prod_{l=1}^{k_i} \frac{\lambda e_i}{\mu_i(l)} P_i(0), \tag{2}$$

$$P_i(0) = \left[1 + \sum_{k_i=1}^{\infty} \prod_{l=1}^{k_i} \frac{\lambda e_i}{\mu_i(l)} \right]^{-1}, \quad i = \overline{1,n}. \tag{3}$$

Here values e_i, $i = \overline{1,n}$, are roots of linear equations system:

$$e_j = p_{0j} + \sum_{i=1}^{n} e_i p_{ij}, \quad j = \overline{1,n}. \tag{4}$$

Thus, in steady state mode the number of messages in the network systems at each fixed time instant is independent random variables.

In case when systems of the Jackson open network are multilinear, i.e. QS S_i consists of m_i identical service lines, and service rate of messages equals μ_i, $i = \overline{1,n}$, for each line, steady probabilities of the network states also look like (1), thus

$$P_i(k_i) = \begin{cases} \frac{\rho_i^{k_i}}{k_i!} P_i(0), k_i \le m_i, \\ \frac{\rho_i^{k_i}}{m_i! m_i^{k_i - m_i}} P_i(0), k_i > m_i, \end{cases}$$

$$P_i(0) = \left[\sum_{j=0}^{m_i} \frac{\rho_i^j}{j!} + \frac{\rho_i^{m_i+1}}{m_i!(m_i - \rho_i)} \right]^{-1},$$

here $\rho_i = \dfrac{e_i \lambda}{\mu_i}$, $i = \overline{1,n}$.

Knowing values of network states probabilities, it is possible to calculate various probability characteristics of the QN. For example, average number of messages N_{i0} in

system S_i queue, average number of messages N_i in system S_i (in queue and on service) and average time of stay τ_i in system S_i are finding by

$$N_{i0} = P_i(m_i)\frac{\rho_i}{(m_i - \rho_i)^2}, \quad i = \overline{1,n},$$

$$N_i = N_{i0} + \sum_{k_i=1}^{m_i-1} k_i P_i(k_i) + \frac{P_i(m_i)}{1 - \frac{\rho_i}{m_i}}, \quad i = \overline{1,n},$$

$$\tau_i = \frac{N_i}{\rho_i \mu_i}, \quad i = \overline{1,n},$$

where $\sum\limits_{k_i=1}^{m_i-1} k_i P_i(k_i) = 0$ when $m_i = 1$.

Proceeding from theoretical results follows that calculation of state probability using (1) can be easy parallelized, for example, each factor $P_i(k_i)$, $i = \overline{1,n}$, can be computed by separate processor irrespectively of other factors. Probability computation process $P_i(0)$, $i = \overline{1,n}$, can also be parallelized, for example, each item of the infinite sum in (3) can be evaluated irrespective of other items on different processors of a computing cluster.

Solution of linear equation system by computing clusters is well studied problem. The matrix of the equation system (4) is rarefied since the matrix of probabilities of message transitions between QS is such. So the problem of values e_i, $i = \overline{1,n}$, finding is reduced to usage of one of well studied methods of the solution of the equation systems multisequencing [4].

3. COSTS OPTIMIZATION IN OPEN JACKSON NETWORK

Let's consider the problem of costs minimization of the QN by the number of service lines in the queueing systems. We introduce next designations: d_i – cost of one message in i-th QS and E_i – cost of one service line in i-th QS, $i = \overline{1,n}$. So the network optimization problem is

$$\begin{cases} W(m) = W(m_1, m_2, ..., m_n) = \sum\limits_{i=1}^{n} (d_i N_i(m_i) + E_i m_i) \to \min\limits_{m_1, m_2, ..., m_n}, \\ m_i \le a_i, \ i = \overline{1,n}, \end{cases} \tag{5}$$

where a_i – given numbers in which limits the number of service lines in i-th QS can change, $i = \overline{1,n}$.

The problem (5) is integer programming problem, it could be solved by exhaustive search mathod. However even for the networks of rather small dimension with number of the systems in the network equal, for example 20, and number of service lines in systems equal to 10, the number of such searches will be 10^{20} and the solution of the problem (5) on one processor could occupy some days.

The MPI technology was applied for parallel computing to solve the optimization problem (5). Software based on programming library PETSc [3] was developed. Open source library PETSc, distributing by GPL license, was developed taking into account usage in the big scalable applications. PETSc includes a huge set of subroutines for parallel solving of linear and non-linear equations that are easily integrated into applications written in programming languages C, C++, Fortran and Python. Also PETSc provides set of mechanisms used for development of parallel programs, for instance creation of parallel matrixes and vectors, allowing to lead a data interchange and to carry out various calculations.

The criterion in optimization problem (5) can be presented in a matrix form:

$$\mathbf{W}\left(m^{(l)}\right)_L = \mathbf{M}\left(m^{(l)}\right)_{L\times 2n} \times \mathbf{E}_{2n} \to \min_{m^{(l)},\, l=\overline{1,L}},$$

where $\mathbf{E}_{2n} = (d_1,\, ...,\, d_n,\, E_1,\, ...,\, E_n)^T$,

$$\mathbf{M}\left(m^{(l)}\right)_{L\times 2n} = \left\| \begin{array}{cccccc} m_1^{(1)} & ... & m_n^{(1)} & N_1\left(m^{(1)}\right) & ... & N_n\left(m^{(1)}\right) \\ m_1^{(2)} & ... & m_n^{(2)} & N_1\left(m^{(2)}\right) & ... & N_n\left(m^{(2)}\right) \\ & & & ... & & \\ m_1^{(L)} & ... & m_n^{(L)} & N_1\left(m^{(L)}\right) & & N_n\left(m^{(L)}\right) \end{array} \right\|,$$

where $m^{(l)} = \left(m_1^{(l)},\, m_2^{(l)}, ...,\, m_n^{(l)}\right)$, $m_i^{(l)}$ – number of service lines in i-th system in l-th search, $L = \prod\limits_{i=1}^{n} a_i$ – number of all possible searches, $l = \overline{1,L}$, $i = \overline{1,n}$.

The process of problem (5) solving with library PETSc is reduced to creation of parallelized matrix $\mathbf{M}\left(m^{(l)}\right)_{L\times 2n}$ arranged line by line on all processors, and further multiplication of it by parallelized vector $\mathbf{E}_{2n}$. The resulting vector $\mathbf{W}\left(m^{(l)}\right)_L$ will be also parallelized. The minimum of cost criterion is a minimum element of vector $\mathbf{W}\left(m^{(l)}\right)_L$ with number l^*, and the solution of problem (5) is the first n elements of a line with number l^* in matrix $\mathbf{M}\left(m^{(l)}\right)_{L\times 2n}$.

4. PROGRAM IMPLEMENTATION

From implementation point of view, the developed software is a bunch of several utilities, each one is dedicated to one task: utility of intermediate and auxiliary matrixes creation, utility for intermediate calculations obtaining (for computing solution of set of linear algebraic equations on a cluster), utility for calculation of QN's characteristics on a cluster, optimization problem utility and also set of utilities for vectors and matrixes visualization.

Specified partition into set independent subroutines allowed to make process of characteristics finding as transparent as possible for end user. Cluster resources are used more effectively since temporary results are not computing several times and could be taken from some intermediate storage. The only user interface with developed utilities is command line, all parameters are specified by command line arguments.

The given set of developed utilities could be used not only on a supercomputer but on regular computers driven by UNIX or Linux operating system with PETSc installed, for example when dimension of task being solving is rather insignificant.

5. THE EXAMPLE

Let's consider queueing network consisting of $n = 5$ QS, the number of service lines in each system can vary from 1 to 10, so the number of possible searches equal $L = 160807$. Probabilities of messages transition between systems are equal $p_{ii+1} = 1,\ i = \overline{0,\, n-1}$, $p_{n0} = 1$, all remaining probabilities in transition matrix are equal to zero. The vector of cost coefficients is $E_{\mathrm{2n}} = (22,\, 31,\, 11,\, 21,\, 13,\, 15,\, 21,\, 43,\, 54,\, 32)^T$.

Solution of optimization problem (5) took about 53 mins on regular AMD64 Linux-based machine in single-processor mode (1 core was used). It took about 17 mins on the same machine with 2 cores dedicated. Obtained results are: $W\left(m^{(l^*)}\right) = 198.275$, $m^{(l^*)} = (2, 1, 2, 2, 3)$, $N\left(m^{(l^*)}\right) = (0.07, 0.953, 0.052, 0.126, 0.009)$.

REFERENCES

1. *Ivnitsky V.A.* Theory of queueing networks, Fizmatlit, Moscow, 2004.
2. *Matalytsky M.A., Tikhonenko O.M. and Pankov A.V.* Queuing theory and its application, GrSU, Grodno, 2008.
3. http://www.mcs.anl.gov/petsc/petsc-as/
4. *Toporkov V.V.* Models of the distributed calculations, Fizmatlit, Moscow, 2004.

СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ СОСТОЯНИЙ СИСТЕМЫ С ДВУМЯ ПРИОРИТЕТАМИ И РЕЗЕРВИРОВАНИЕМ КАНАЛОВ

А. Печинкин

Институт проблем информатики РАН
Москва, Россия
apechinkin@ipiran.ru

Рассматривается марковская модель двухприоритетной системы, обобщающей систему с резервированием каналов пучка, марковским входящим потоком и различными распределениями фазового типа времен обслуживания заявок каждого приоритета. Резервирование каналов пучка означает наличие некоторого числа каналов, которые могут занимать только приоритетные заявки. Показано, как можно вычислить стационарные вероятности состояний этой марковской модели.

Ключевые слова: система массового обслуживания; резервирование каналов пучка; стационарные вероятности состояний.

1. ВВЕДЕНИЕ

Одной из возможных схем обслуживания заявок нескольких приоритетов является система с резервированием каналов пучка (СРКП), в которой для заявок i-го приоритета имеется свой порог n_i, и заявки данного приоритета принимаются к обслуживанию только в том случае, когда число занятых каналов меньше n_i. Такая схема обслуживания заявок, получившая в литературе название системы с резервированием каналов пучка (СРКП), впервые была описана P. J. Burke (фирма Bell, США) в 1961 г. для обслуживания двухприоритетного пучка и впоследствии использована в ряде систем связи, в частности, в системе RITA (франция) [1], [2]. Различные модификации СРКП исследовались в работах [1], [3], [4], [5], [6], [7], [8], [9] и других.

В [10] была предложена методика расчета СРКП с произвольным числом приоритетов, пуассоновскими входящими потоками и одинаковым экспоненциальным обслуживанием заявок всех приоритетов. Эта методика позволила получить в [11] легко реализуемые алгоритмы вычисления основных стационарных характеристик двухприоритетной СРКП с марковским входящим потоком и одинаковым экспоненциальным обслуживанием заявок всех приоритетов. Однако предположения об экспоненциальности и одинаковой распределенности времен обслуживания заявок различных приоритетов существенно ограничивают сферу применения результатов [10] и [11]. В настоящей работе предложена свободная от этих недостатков марковская модель, обобщающая двухприоритетную СРКП. Показано, как для этой

модели можно вычислить стационарные распределения очередей заявок разных приоритетов.

2. ОПИСАНИЕ МОДЕЛИ

Рассмотрим систему массового обслуживания марковского типа с входящим потоком заявок двух типов (далее заявки первого типа будем называть приоритетными, а второго — неприоритетными) и относительным приоритетом, описываемую следующим образом.

Имеются неотрицательные целые числа $n_0 \geq 1$, $n_1 \geq n_0$ и $n_2 \geq 0$. Эти числа интерпретируются так: n_0 — число приборов, доступных всем заявкам (иными словами, если занято менее n_0 приборов, то любая поступающая заявка немедленно начинает обслуживаться); n_1 — максимальное суммарное число неприоритетных заявок, находящихся на приборах, и приоритетных заявок, находящихся в системе; n_2 — число мест ожидания для неприоритетных заявок, т.е. максимальное число неприоритетных заявок, которые могут находиться в системе дополнительно к заявкам на приборах.

Процесс поступления и обслуживания заявок определяется матрицами $\Lambda_n^{(u)}$, $n = \overline{0, n_1}$, $u = 1, 2$, M_n, $n = \overline{1, n_1}$, N_n, $n = \overline{0, n_1}$, и Ω (размеры матриц определяются далее и задаются числами I_n, $n = \overline{0, n_1}$, которые будем называть числами фаз процесса поступления-обслуживания заявок в слоях n, или просто числами фаз) и протекает таким образом.

Если в системе находится n, $n = \overline{0, n_0 - 1}$, заявок (далее будем говорить, что процесс поступления-обслуживания находится в слое n) и фаза равна i, $i = \overline{1, I_n}$, то с интенсивностью $(\Lambda_n^{(u)})_{ij}$, $u = 1, 2$, $j = \overline{1, I_{n+1}}$, в систему поступает новая заявка u-го типа, которая тут же начинает обслуживаться, и фаза становится равной j. Кроме того, с интенсивностью $(N_n)_{ij}$, $j = \overline{1, I_n}$, $j \neq i$, в систему не поступают заявки и не заканчивается обслуживание ни одной из n заявок, но фаза становится равной j. Наконец, если дополнительно $n \geq 1$, то с интенсивностью $(M_n)_{ij}$, $j = \overline{1, I_{n-1}}$, заканчивается обслуживание одной из n заявок, процесс поступления-обслуживания переходит в слой $n - 1$ и фаза становится равной j.

При $n = \overline{n_0 + 1, n_1 - 1}$ будем говорить, что процесс поступления-обслуживания находится в слое n в том случае, когда суммарное число неприоритетных заявок, находящихся на приборах, и приоритетных заявок, находящихся в системе, равно n и, возможно, имеется еще какое-то число (но не более n_2) неприоритетных заявок в очереди. При этом если фаза процесса поступления-обслуживания равна i, $i = \overline{1, I_n}$, то с интенсивностью $(\Lambda_n^{(1)})_{ij}$, $j = \overline{1, I_{n+1}}$, в систему поступает приоритетная заявка, причем, как и прежде, процесс поступления-обслуживания переходит в слой $n+1$ и фаза становится равной j. С интенсивностью $(N_n)_{ij}$, $j = \overline{1, I_n}$, $j \neq i$, в систему не поступают заявки и не заканчивается обслуживание заявок, но с i-й на j-ю изменяется фаза, а с интенсивностью $(M_n)_{ij}$, $j = \overline{1, I_{n-1}}$, заканчивается обслуживание одной из заявок и фаза становится равной j. Однако если в систему поступает неприоритетная заявка (с интенсивностью $(\Lambda_n^{(2)})_{ij}$, $j = \overline{1, I_n}$), то слой

n не меняется, а фаза становится равной j. Поступающая неприоритетная заявка либо становится в очередь неприоритетных заявок (если там имеются свободные места), либо теряется.

Единственным отличием случая $n = n_1$ от предыдущего является то, что поступающая приоритетная заявка теряется, изменяя фазу в том же слое. Естественно, матрица $\Lambda_{n_1}^{(1)}$ в этом случае является квадратной порядка I_{n_1}.

Обратимся к последнему случаю: $n = n_0$. Здесь возможны два варианта, связанные с окончанием обслуживания. Первый вариант появляется, если в системе в очереди отсутствуют неприоритетные заявки. Тогда, как и раньше, с интенсивностью $(M_{n_0})_{ij}$, $i = \overline{1, I_{n_0}}$, $j = \overline{1, I_{n_0-1}}$, заканчивается обслуживание одной из заявок, процесс поступления-обслуживания переходит в слой $n_0 - 1$ и фаза становится равной j. Прежний смысл имеют также матрицы $\Lambda_{n_0}^{(u)}$, $u = 1, 2$, и N_{n_0}. Если же в системе в очереди имеются неприоритетные заявки (второй вариант), то с интенсивностью $(M_{n_0})_{ij}$, $i = \overline{1, I_{n_0}}$, $j = \overline{1, I_{n_0-1}}$, заканчивается обслуживание заявки, но на освободившийся прибор мгновенно поступает заявка из очереди неприоритетных заявок, с вероятностью Ω_{jl}, $l = \overline{1, I_{n_0}}$, меняя фазу с j-й на l-ю и возвращая процесс поступления-обслуживания в слой n_0. Матрицы $\Lambda_{n_0}^{(u)}$ имеют прежний смысл.

В тех случаях, когда суммирование возможно, положим:

$$N_n^{(0)} = N_n + \Lambda_n^{(1)} + \Lambda_n^{(2)}; \quad N_n^{(1)} = N_n + \Lambda_n^{(2)}; \quad N_n^{(2)} = N_n + \Lambda_n^{(1)}.$$

3. СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ СОСТОЯНИЙ

Пусть $\nu(t)$ обозначает слой, в котором находится процесс поступления-обслуживания в момент t. Предположим, что в начальный момент значение процесса $\nu(t)$ равнялось n, $n = \overline{n_0, n_1}$, отсутствовала очередь неприоритетных заявок и фаза была i. Введем матрицы $F_k(n)$, $k = \overline{0, n_2 - 1}$, и $\tilde{F}_k(n)$, $k = \overline{0, n_2}$. Элементы $(F_k(n))_{ij}$ и $(\tilde{F}_k(n))_{ij}$, $i = \overline{1, I_n}$, $j = \overline{1, I_{n-1}}$, матриц $F_k(n)$ и $\tilde{F}_k(n)$ представляют собой вероятности того, что непосредственно после момента, когда впервые значение процесса $\nu(t)$ станет равным $n - 1$ (в случае $n = n_0$ таким моментом может быть также момент, когда на прибор впервые поступит неприоритетная заявка из очереди неприоритетных заявок, однако эта неприоритетная заявка пока еще не учитывается как поступившая на прибор), фаза будет j и в системе в очереди неприоритетных заявок будет находиться ровно k и, соответственно по крайней мере k, заявок. Следующие соотношения позволяют рекуррентно по n и k, начиная с $n = n_1$ и $k = 0$, вычислять матрицы $F_k(n)$ и $\tilde{F}_k(n)$:

$$F_0(n_1) = (-N_{n_1}^{(2)})^{-1} M_{n_1}; \tag{1}$$

$$\tilde{F}_0(n_1) = (-N_{n_1}^{(0)})^{-1} M_{n_1}; \tag{2}$$

$$F_k(n_1) = (-N_{n_1}^{(2)})^{-1} \Lambda_{n_1}^{(2)} F_{k-1}(n_1), \quad k = \overline{1, n_2 - 1}; \tag{3}$$

$$\tilde{F}_k(n_1) = (-N_{n_1}^{(2)})^{-1} \Lambda_{n_1}^{(2)} \tilde{F}_{k-1}(n_1), \quad k = \overline{1, n_2}; \tag{4}$$

$$F_0(n) = -N_n^{-1}\left[\Lambda_n^{(1)} F_0(n+1) F_0(n) + M_n\right], \quad n = \overline{n_0, n_1 - 1}; \tag{5}$$

$$\tilde{F}_0(n) = (-N_n^{(1)})^{-1}\left[\Lambda_n^{(1)} \tilde{F}_0(n+1) \tilde{F}_0(n) + M_n\right], \quad n = \overline{n_0, n_1 - 1}; \tag{6}$$

$$F_k(n) = -N_n^{-1}\left[\Lambda_n^{(1)} \sum_{i=0}^{k} F_i(n+1) F_{k-i}(n) + \right.$$

$$\left. + \Lambda_n^{(2)} F_{k-1}(n)\right], \quad k = \overline{1, n_2 - 1}, \quad n = \overline{n_0, n_1 - 1}; \tag{7}$$

$$\tilde{F}_k(n) = -N_n^{-1}\left[\Lambda_n^{(1)} \left(\sum_{i=0}^{k-1} F_i(n+1) \tilde{F}_{k-i}(n) + \tilde{F}_k(n+1) \tilde{F}_0(n)\right) + \right.$$

$$\left. + \Lambda_n^{(2)} \tilde{F}_{k-1}(n)\right], \quad k = \overline{1, n_2}, \quad n = \overline{n_0, n_1 - 1}. \tag{8}$$

Положим $F_k = F_k(n_0)$, $k = \overline{0, n_2 - 1}$, и $\tilde{F}_k = \tilde{F}_k(n_0)$, $k = \overline{0, n_2}$.

Вложенную цепь Маркова для данной системы можно ввести различными способами. Удобно определить ее таким образом. Рассмотрим следующие моменты: моменты освобождения каких-либо приборов или поступления заявок в систему в случае, когда перед этими моментами было занято менее n_0 приборов; моменты освобождения каких-либо приборов в случае, когда перед этими моментами было занято ровно n_0 приборов. Последовательность таких моментов обозначим через $\{\tau_l, \quad l \geq 1\}$. Отметим, что непосредственно после момента второго типа либо в системе остается ровно $n_0 - 1$ заявок (все они обслуживаются на приборах), либо на прибор поступает неприоритетная заявка из очереди. Множество состояний вложенной цепи Маркова представляет собой пару (i, m), $i = \overline{1, I_m}$, $m = \overline{0, n_0 + n_2 - 1}$. Здесь i — фаза (процесса поступления-обслуживания) соответствующего слоя. Более хитро определяется второй компонент m множества состояний. А именно, если $m = \overline{0, n_0 - 1}$, то, как обычно, m — число заявок в системе. Однако, если $m = \overline{n_0, n_0 + n_2 - 1}$, то число занятых приборов равно n_0, а $m - n_0$ — число неприоритетных заявок в очереди. Саму цепь Маркова образуют фаза и число занятых приборов (номер слоя) или число занятых приборов (номер слоя) плюс число неприоритетных заявок в очереди непосредственно после моментов τ_l.

Матрицу переходных вероятностей вложенной цепи Маркова будем обозначать через $P = (P_{m_1 m_2})$, $m_1, m_2 = \overline{0, n_0 + n_2 - 1}$. Отметим, что в силу принятого нами соглашения элементы $P_{m_1 m_2}$ матрицы P сами являются матрицами размеров, определяемых числами фаз соответствующих слоев. Эти матрицы задаются формулами:

$$P_{m,m-1} = -N_m^{-1} M_m, \quad m = \overline{1, n_0 - 1};$$

$$P_{m,m+1} = -N_m^{-1}(\Lambda_m^{(1)} + \Lambda_m^{(2)}), \quad m = \overline{0, n_0 - 1};$$

$$P_{m_1 m_2} = F_{m_2 - m_1 + 1}\Omega, \quad m_1 = \overline{n_0, n_0 + n_2 - 1}, \quad m_2 = \overline{m_1 - 1, n_0 + n_2 - 2}, \quad m_2 \neq n_0 - 1;$$

$$P_{n_0, n_0 - 1} = F_0;$$

$$P_{m,n_0+n_2-1} = \tilde{F}_{n_0+n_2-m}\Omega, \quad m = \overline{n_0, n_0+n_2-1}.$$

Остальные матрицы $P_{m_1 m_2}$ являются нулевыми.

Обозначим через $\vec{p}^*$ вектор-строку стационарных вероятностей вложенной цепи Маркова. Естественно, координаты $\vec{p}^*_m$, $m = \overline{0, n_0+n_2-1}$, вектора $\vec{p}^*$ сами являются вектор-строками с координатами $(\vec{p}^*_m)_i$, где индекс i означает фазу.

Вектор $\vec{p}^*$ удовлетворяет системе уравнений равновесия (СУР)

$$\vec{p}^* = \vec{p}^* P$$

с условием нормировки

$$\sum_{m=0}^{n_0+n_2-1} \vec{p}^*_m \vec{1} = 1.$$

Здесь через $\vec{1}$ обозначен вектор-столбец $(1, \ldots, 1)^T$, размерность которого определяется размером умножаемой на него слева матрицы. Для решения СУР удобно применить алгоритм, основанный на последовательном упрощении цепи Маркова с помощью исключения множеств состояний (см. [12], стр. 22).

Предположим теперь, что в начальный момент значение процесса $\nu(t)$ равнялось n, $n = \overline{n_0, n_1}$, отсутствовала очередь неприоритетных заявок и фаза была i. Обозначим через $S_k(n)$, $k = \overline{n, n_1}$, матрицу, элементом $(S_k(n))_{ij}$, $j = \overline{1, I_k}$, которой является среднее время, проведенное системой в состояниях, когда значение процесса $\nu(t)$ равнялось k и фаза была j, до того момента, когда впервые значение процесса $\nu(t)$ стало равным $n-1$ (в случае $n = n_0$ таким моментом может быть также момент, когда на прибор впервые поступит неприоритетная заявка из очереди неприоритетных заявок), а через $\vec{t}_k(n)$, $k = \overline{0, n_2-1}$, и $\vec{T}_k(n)$, $k = \overline{0, n_2}$, — вектор-столбцы, координатами $(\vec{t}_k(n))_i$ и $(\vec{T}_k(n))_i$, $i = \overline{1, I_n}$, которых являются средние времена, проведенные системой в состояниях, когда в очереди неприоритетных заявок имелось ровно k и, по крайней мере k, заявок (без учета фазы), до того момента, когда впервые значение процесса $\nu(t)$ стало равным $n-1$ (с прежним замечанием относительно случая $n = n_0$). Матрицы $S_k(n)$ и векторы $\vec{t}_k(n)$ и $\vec{T}_k(n)$ удовлетворяют уравнениям, аналогичным (1)–(3), и также легко вычисляются рекуррентно.

Наконец, обозначим через $\vec{p}_n$, $n = \overline{0, n_1}$, — вектор-строку с координатами $(\vec{p}_n)_i$, $i = \overline{1, I_n}$, представляющими собой стационарные вероятности того, что значение процесса $\nu(t)$ равно n и фаза равна i, а через $p^{(2)}_n$, $n = \overline{0, n_2}$, — стационарную вероятность того, что в системе в очереди неприоритетных заявок имеется n заявок (фаза не учитывается). Тогда:

$$\vec{p}_n = -\frac{1}{\overline{m}} \vec{p}^*_n N_n^{-1}, \quad n = \overline{0, n_0-1}; \qquad \vec{p}_n = \frac{1}{\overline{m}} \sum_{k=n_0}^{n_0+n_2-1} \vec{p}^*_k S_n, \quad n = \overline{n_0, n_1};$$

$$p^{(2)}_0 = \frac{1}{\overline{m}} \Big(-\sum_{k=0}^{n_0-1} \vec{p}^*_k N_k^{-1} \vec{1} + \vec{p}^*_{n_0} \vec{t}_0 \Big); \qquad p^{(2)}_n = \frac{1}{\overline{m}} \sum_{k=0}^{n} \vec{p}^*_{n_0+k} \vec{t}_{n-k}, \quad n = \overline{1, n_2-1};$$

$$p_{n_2}^{(2)} = \frac{1}{\overline{m}} \sum_{k=0}^{n_2-1} \vec{p}^{\,*}_{n_0+k} \vec{T}_{n_2-k},$$

где:

$$\overline{m} = \sum_{n=0}^{n_0+n_2-1} \vec{p}^{\,*}_n \vec{m}_n;$$

$$\vec{m}_n = -N_n^{-1}\vec{1}, \quad n = \overline{0, n_0-1}; \qquad \vec{m}_n = \vec{T}_0, \quad n = \overline{n_0, n_0+n_2-1}.$$

ЛИТЕРАТУРА

1. *Grandjean C. H.* Traffic Calculations in Suturation Routing with Priorities // Electr. Commun. 1974. V. 49. № 1, P. 72–79.
2. *Людвиг Г., Рой Р.* Ограничения для сетей с волновым поиском сетей // Тр. инта инженеров электроники и радиоэлектроники. 1977. Т. 65. № 9. С. 154–165.
3. *Weber J. H.* Some Traffic Characteristics of Communications Networks with Automatic Alternate Routing // Bell System Techn. J. 1962. March. P. 1201–1247.
4. *Weber J. H.* Simulation Study of Routing and Control in Communications Networks // Bell System Techn. J. 1964. Nov. P. 2639–2676.
5. *Grandjean C. H.* Call Routing Strategies in Telecommunications Networks // Electr. Commun. 1967. V. 42, № 3. P. 380–391.
6. *Джейсуол Н.* Очереди с приоритетами. М.: Мир, 1973.
7. *Esoqbue A. O., Singh A. J.* A Stochastic Model for a Optimal Priority Bed Distribution in a Hospital // Oper. Res. 1976. № 24. P. 884–889.
8. *Otterman J.* Grande of Service Direct Traffic mixed with Store-and-Forward Traffic // Bell System Techn. J. 1962. Apr. P. 1415–1437.
9. *Liu F. K.* A Combined Delay and Loss System with Priority // ICC. 1973. V. 39. № 7. P. 39-7–39-13.
10. *Печинкин А. В., Федоров В. М.* Методика расчета многоканальной системы приоритетного обслуживания с резервированием каналов // Системное моделирование. Вып. 15. Новосибирск: ВЦ СО АН СССР, 1990.
11. *Бурыгин С. В., Глазунов А. С., Печинкин А. В.* Система приоритетного обслуживания с резервированием каналов и марковским входящим потоком // Вестник Российского университета дружбы народов. Сер. Прикладная математика и информатика. 2001. № 1. С. 80–89.
12. *Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S.* Queueing Theory. Utrecht–Boston: VSP, 2004.

О ВРЕМЕНИ ОЖИДАНИЯ ПРИ НЕКОТОРЫХ ДИСЦИПЛИНАХ ОБСЛУЖИВАНИЯ В СИСТЕМЕ С ОТРИЦАТЕЛЬНЫМИ ЗАЯВКАМИ И БУНКЕРОМ

А. Печинкин[1], Р. Разумчик[2,*]

[1] *Институт проблем информатики РАН,*
[2] *Российский Университет Дружбы Народов*
[1,2] *Москва, Россия*
[*] rrazumchik@ieee.org

Рассматривается однолинейная система массового обслуживания с отрицательными заявками и накопителем бесконечной емкости. Отрицательная заявка вытесняет (обычную) заявку из накопителя в бункер также бесконечной емкости. Заявки из накопителя обслуживаются с относительным приоритетом по отношению к заявкам из накопителя. Для двух дисциплин выбора на обслуживание и вытеснения заявок получено в терминах преобразования Лапласа–Стилтьеса распределение времени ожидания начала обслуживания.

Ключевые слова: отрицательные заявки, бункер, время ожидания.

1. ОПИСАНИЕ СИСТЕМЫ

Рассмотрим однолинейную систему массового обслуживания (СМО), в которую поступает пуассоновский поток заявок интенсивности λ. Заявки этого потока будем называть положительными. Для них имеется накопитель неограниченной емкости. Дополнительно в систему поступает пуассоновский поток отрицательных заявок интенсивности λ^-. Отрицательная заявка перемещает одну положительную заявку из накопителя в бункер неограниченной емкости и после этого покидает систему. Если в момент поступления отрицательной заявки в накопителе нет положительных заявок, то, вне зависимости от состояния прибора, отрицательная заявка покидает систему, не оказывая на нее никакого воздействия.

После окончания обслуживания очередной заявки на прибор становится заявка из накопителя или, если накопитель пуст, из бункера. Прерывание обслуживания заявок не допускается. Длительности обслуживания заявок, как из накопителя, так и из бункера, имеют экспоненциальное распределение с одним и тем же параметром μ. Далее будем предполагать, что выполнено необходимое и достаточное

условие $\rho = \lambda/\mu < 1$ существования стационарного режима функционирования СМО.

Описанная выше система с отрицательными заявками, которые не «убивают» находящиеся в системе заявки, а лишь откладывают их обслуживание на случайное время, вытесняя в другую очередь, была введена в [1], где были получены результаты, связанные со стационарными распределениями (совместным и маргинальными) чисел заявок в системе.

В настоящей работе находится распределение времени ожидания начала обслуживания в рассматриваемой СМО для следующих дисциплин выбора из накопителя и бункера на обслуживание и вытеснения из накопителя в бункер:

- при поступлении отрицательной заявки вытесняется последняя заявка из очереди в накопителе, а в момент окончания обслуживания заявки на приборе на обслуживание выбирается первая заявка из очереди в накопителе или, если накопитель пуст, первая заявка из очереди в бункере (дисциплина $LAST$-$FIFO$-$FIFO$);
- при поступлении отрицательной заявки вытесняется первая заявка из очереди в накопителе, а в момент окончания обслуживания заявки на приборе на обслуживание выбирается последняя заявка из очереди в накопителе или, если накопитель пуст, первая заявка из очереди в бункере (дисциплина $FIRST$-$LIFO$-$FIFO$);

2. ВСПОМОГАТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Пусть $\{p_{km},\ \ m \geq 0,\ \ k \geq u(m)\}$ (здесь и далее через $u(x)$ обозначается функция Хевисайда) — стационарное распределение вероятностей того, что в накопителе и на приборе находится k заявок, а в бункере — m заявок. Тогда, используя вид двойной производящей функции (ПФ) $P(u,v) = p_{00} + \sum\limits_{k=1}^{\infty}\sum\limits_{m=0}^{\infty} p_{km}u^k v^m$, найденный в [1], можно показать, что

$$\sum_{m=0}^{\infty} p_{km}v^m = (1-\rho)\left(1 - \frac{\lambda^-}{\lambda u_2 - \mu}\right)\frac{1}{u_1^k}\,, \quad k \geq 1, \tag{1}$$

где

$$u_{1,2} = u_{1,2}(v) = \frac{\lambda + \mu + \lambda^- \pm \sqrt{(\lambda+\mu+\lambda^-)^2 - 4\lambda(\mu + \lambda^- v)}}{2\lambda}\,.$$

Подставляя $v = 1$ в (1), легко найти вид вероятностей $\{p_{i,\cdot}\ \ i \geq 0\}$ того, что в накопителе и на приборе находятся i заявок:

$$p_{i,\cdot} = \begin{cases} 1-\rho, & i = 0, \\ \rho(1-q)q^{i-1}, & i \geq 1, \quad q = \dfrac{\lambda}{\mu + \lambda^-}\,. \end{cases} \tag{2}$$

Рассмотрим теперь период занятости (ПЗ) системы $M/M/1/\infty$ с входящим потоком интенсивности λ и интенсивностью обслуживания b. Обозначим через $G(x;b)$ функцию распределения (ФР) ПЗ этой системы, а через $\gamma(s;b) = \int\limits_0^\infty e^{-sx} dG(x;b)$ — преобразование Лапласа–Стилтьеса (ПЛС) ФР $G(x;b)$. Тогда (см., например [2, 3])

$$\gamma(s;b) = \frac{s+\lambda+b-\sqrt{(s+\lambda+b)^2-4\lambda b}}{2\lambda}\,.$$

Обозначим через $G(x,k;b)$ вероятность того, что ПЗ этой системы будет меньше x и на нем обслужится ровно k заявок. Двойное преобразование (ПЛС по x и ПФ по k) $\gamma(s,z;b) = \sum\limits_{k=1}^\infty z^k \int\limits_0^\infty e^{-sx} dG(x,k;b)$ задается формулой

$$\gamma(s,z;b) = \frac{s+\lambda+b-\sqrt{(s+\lambda+b)^2-4\lambda z b}}{2\lambda}\,.$$

Наконец, через $G^*(x)$ обозначим ФР ПЗ СМО $M/M/1/\infty$ с параметрами $\lambda^* = \lambda\mu/(\mu+\lambda^-)$ и μ. Отметим, что $G^*(x)$ представляет собой также ФР ПЗ СМО $M/M/1/\infty$ с параметрами λ и $(\mu+\lambda^-)$, открываемого заявкой экспоненциальной с параметром μ длины. ПЛС ФР $G^*(x)$ имеет вид

$$\gamma^*(s) = \frac{\mu}{\mu+s+\lambda[1-\gamma(s;\mu+\lambda^-)]}\,.$$

3. ДИСЦИПЛИНА *LAST-FIFO-FIFO*

Найдем стационарное распределение времени ожидания начала обслуживания поступившей в систему заявки при дисциплине *LAST-FIFO-FIFO*.

Обозначим через $F(x)$ стационарную вероятность того, что попавшая в накопитель заявка поступит из накопителя на прибор, причем за время меньше x, при условии, что за это время не будет вытеснена («убита») ни одна заявка. Тогда

$$F(x) = \frac{1}{\rho}\sum_{i=1}^\infty p_{i,\cdot}\, H_i(x;\mu)$$

(здесь и далее $H_i(x;\mu)$ — ФР Эрланга с i фазами и интенсивностью обслуживания μ). С учетом (2) вероятность $F(x)$ в терминах ПЛС задается выражением

$$\psi(s) = (1-q)\sum_{i=1}^\infty q^{i-1}\left(\frac{\mu}{s+\mu}\right)^i = \frac{\mu(1-q)}{s+\mu(1-q)}\,,$$

обращая которое, имеем:

$$f(x) = F'(x) = \mu(1-q)e^{-\mu(1-q)x}, \quad x>0.$$

Вероятность того, что заявка поступит из накопителя в бункер, причем за время меньше x, при условии, что за время x не будет обслужена ни одна заявка, имеет ФР $G(x;\lambda^-)$ с ПЛС $\gamma(s;\lambda^-)$.

Таким образом, стационарная вероятность $V_{nak}(x)$ того, что заявка поступит из накопителя сразу же на прибор, причем за время меньше x, имеет ПЛС

$$\varphi_{nak}(s) = p_0 + \rho \int\limits_0^\infty e^{-sx}[1-G(x;\lambda^-)]f(x)\,dx = 1-\rho + \frac{\lambda(1-q)[1-\gamma(s+\mu(1-q);\lambda^-)]}{s+\mu(1-q)}.$$

Перейдем к вычислению стационарной вероятности $V_{bun}(x)$ того, что заявка поступит из накопителя в бункер, а затем из бункера на прибор, причем за время меньше x.

Обозначим через $F_n(x,k,l)$, $n \geq 0$, $k = \overline{0,n}$, $l \geq 1$, вероятность того, что выделенная заявка попадет из накопителя в бункер, причем за время меньше x, и за это время в бункер поступит еще l заявок , $l \geq 1$, а в момент перехода в бункер этой заявки в накопителе будет k заявок, при условии, что в момент поступления выделенной заявки в систему в накопителе было n заявок. Тогда, учитывая, что $g(x,l;b) = G'_x(x,k;b)$, имеем

$$f_n(x,k,l) = F'_n(x,k,l) = \frac{(\mu x)^{n-k}}{(n-k)!}e^{-\mu x}g(x,l;\lambda^-), \quad n \geq 0, \quad k = \overline{0,n}, \quad l \geq 1.$$

Поскольку после перехода выделенной заявки в буфер она до поступления на прибор будет ждать время обслуживания заявки на приборе, заявок в бункере и всех их «неубитых» потомков, а также время обслуживания «неубитых» заявок в накопителе и их «неубитых» потомков. Поэтому ПЛС стационарной вероятности $V_{bun}(x)$ определяется формулой

$$\varphi_{bun}(s) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} p_{n+1,m} \int\limits_0^\infty e^{-sx} \sum_{k=0}^{n}\sum_{l=1}^{\infty} f_n(x,k,l)(\gamma^*(s))^{l+m}(\gamma(s;\mu+\lambda^-))^k dx,$$

которая, после упрощения с учетом (1), принимает следующий вид:

$$\varphi_{bun}(s) = \frac{(1-\rho)(\lambda u_2(\gamma^*(s)) - \mu - \lambda^-)\gamma\left(s+\mu-\frac{\mu}{u_1(\gamma^*(s))}, \gamma^*(s);\lambda^-\right)}{[\lambda u_2(\gamma^*(s)) - \mu][u_1(\gamma^*(s)) - \gamma(s;\mu+\lambda^-)]}.$$

Стационарное распределение общего времени ожидания начала обслуживания поступившей в систему заявки $V(x)$ находится в терминах в терминах ПЛС $\varphi(s) = \int\limits_0^\infty e^{-sx}dV(x)$ по формуле:

$$\varphi(s) = \varphi_{nak}(s) + \varphi_{bun}(s).$$

Дифференцируя последнюю формулу соответствующее число раз, нетрудно получить выражения для моментов любого порядка стационарного распределения времени ожидания начала обслуживания заявки. В частности, среднее общее время ожидания начала обслуживания равно

$$-\varphi'(0) = \frac{\lambda}{\mu(\mu-\lambda)}. \tag{3}$$

4. ДИСЦИПЛИНА *FIRST-LIFO-FIFO*

Найдем теперь стационарное распределение времени ожидания начала обслуживания поступившей в систему заявки при дисциплине *FIRST-LIFO-FIFO*.

Вводя стационарную вероятность $F(x)$ того, что заявка поступит из накопителя в бункер, причем за время меньше x, при условии, что за это время не обслужится ни одна заявка, и используя те же самые рассуждения, что и в предыдущем разделе, получаем для нее следующее выражение:

$$F(x) = \frac{1}{\rho}\sum_{i=1}^{\infty} p_{i,\cdot}\, H_i(x;\lambda^-) = 1 - e^{-\lambda^-(1-q)x},$$

а для ПЛС стационарной вероятности $V_{nak}(x)$ того, что заявка поступит из накопителя сразу же на прибор, причем за время меньше x, — выражение

$$\varphi_{nak}(s) = p_0 + \rho\int\limits_0^{\infty} e^{-sx}[1 - F(x)]\, dG(x;\mu) = 1 - \rho + \rho\gamma(s + \lambda^-(1-q);\mu).$$

Перейдем к вычислению стационарной вероятности $V_{bun}(x)$ того, что заявка поступит из накопителя в бункер, а затем из бункера на прибор, причем за время меньше x.

Обозначим через $F(x,k)$, $k \geq 0$, вероятность того, что за время x заявка не поступит из накопителя на прибор и в момент x за ней будет k других заявок, при условии, что за время x не будет убита ни одна заявка. Эта вероятность совпадает с вероятностью того, что ПЗ СМО $M/M/1/\infty$ с параметрами λ и μ не закончится к моменту x и в момент x в этой системе в очереди будет k заявок. Двойное преобразование (преобразование Лапласа по x и ПФ по k) вероятности $F(x,k)$ имеет вид (см., например, [2], гл. 5, § 5.2)

$$\psi(s,z) = \frac{\phi(s,z)}{\mu + s + \lambda - \lambda z}, \qquad \phi(s,z) = \frac{z - \gamma(s;\mu)}{z - \dfrac{\mu}{\mu + s + \lambda - \lambda z}}.$$

Далее, вероятность того, что заявка, попавшая в накопитель, в котором находится еще n заявок, переместится в бункер, причем за время меньше x, при условии, что за это время не обслужится ни одна заявка, равна $H_{n+1}(x;\lambda^-)$.

После перехода выделенной заявки в буфер она до поступления на прибор будет ждать еще время обслуживания заявки на приборе и заявок в бункере (а к моменту перехода выделенной заявки в буфер все n заявок, находившихся в накопителе в момент поступления в систему выделенной, переместятся в бункер) и всех их «неубитых» потомков, а также время обслуживания k «неубитых» заявок в накопителе и их «неубитых» потомков. Значит, ПЛС стационарной вероятности $V_{bun}(x)$ определяется формулой

$$\varphi_{bun}(s) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} p_{k+1,m}\int\limits_0^{\infty} e^{-sx}\sum_{i=0}^{\infty} F(x,i)(\gamma^*(s))^{k+m+1}(\gamma(s;\mu+\lambda^-))^i\, dH_{k+1}(x;\lambda^-),$$

которая после упрощения принимает с учетом (1) следующий вид:

$$\varphi_{bun}(s) = \frac{\lambda^-(1-\rho)\gamma^*(s)(\lambda u_2 - \mu - \lambda^-)}{(\lambda u_2 - \mu)u_1(\gamma^*(s))}\psi\left(s + \lambda^- - \frac{\lambda^-\gamma^*(s)}{u_1(\gamma^*(s))}, \gamma(s; \mu + \lambda^-)\right).$$

Как и в предыдущем разделе, ПЛС стационарного распределения общего времени ожидания начала обслуживания поступившей в систему заявки имеет вид

$$\varphi(s) = \varphi_{nak}(s) + \varphi_{bun}(s),$$

а среднее общее время ожидания определяется формулой (3).

5. ЗАКЛЮЧЕНИЕ

Как показано в предыдущих разделах и как следует из формулы Литтла, средние общие времена ожидания начала обслуживания для поступающей заявки совпадают для обеих рассмотренных дисциплин. Между тем, значения дисперсий времени ожидания различны.

Проведенные численные расчеты показали, что наилучшей (обладающей минимальной дисперсией времени ожидания начала обслуживания заявки) дисциплиной вне зависимости от исходных параметров λ, μ и λ^- является дисциплина $LAST\text{-}FIFO\text{-}FIFO$.

Заметим, что при $\lambda^- \to \infty$ дисперсии времени ожидания обеих дисциплин, как и должно быть, стремятся к дисперсии в системе $M/M/1/\infty$ при отсутствии отрицательных заявок и обслуживании заявок из накопителя в порядке поступления, т.е. к $\rho(2-\rho)/(\mu^2(1-\rho)^2)$.

Для проверки правильности полученных аналитических соотношений была разработана с помощью программных средств GPSS [4] имитационная модель, результаты расчетов по которой практически совпали с результатами численных расчетов на основе полученных аналитических соотношений.

ЛИТЕРАТУРА

1. *Мандзо Р., Касконе Н., Разумчик Р. В.* Экспоненциальная система массового обслуживания с отрицательными заявками и бункером для вытесненных заявок // Автоматика и телемеханика. 2008. V. 9. P. 103–113.
2. *Бочаров П. П., Печинкин А. В.* Теория массового обслуживания. М.: Изд-во РУДН, 1995.
3. *Kleinrock L.* Queueing Systems: Volume I – Theory. New York: Wiley Interscience, 1975.
4. *Бражник А. Н.* Имитационное моделирование: возможности GPSS World. СПб.: Реноме, 2006.

ON APPROXIMATION OF RETRIAL QUEUES WITH VARYING SERVICE RATE

V. Ponomarov

Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
vponomarov@gmail.com

The paper deals with Markov model of retrial queueing system in which service rate depends on queue length. Investigation method is based on an approximation of the input system by the system with truncated state space. The accuracy of such approximation is also discussed.

Keywords: Stochastic system, retrials, steady state.

1. INTRODUCTION

The research of wide class of stochastic systems with repeated calls faces the problem of calculating characteristics of the system with the poisson incoming flow. Markov process that describes the behavior of such system has an infinite state space and the transaction matrix usually does not have special properties that simplify the process of finding Kolmogorov set of equations explicit solution. Described features lead to the fact that only few simple models were examined in details [1].

Usually the problem of computing stationary probabilities for such systems is solved by computation algorithms or recurrent schemas [2]. To apply them system with finite state space is taken into account. Usually it is considered that the queue length does not exceed the value of M. If the request comes to the system when there is no free server and there are M sources of repeated calls already it is lost. It is intuitively obvious that by choosing M big enough we can approximate characteristics of the input system with the predefined accuracy.

In this paper we present the described above technic for the retrial system with Poisson input flow and varying service rate. Its integral characteristics are approximated by characteristics of the system with truncated space state which can be explicitly written in terms of system's parameters. The error of such approximation for different service rate switching policies is also discussed.

2. MARKOV MODELS OF THE INPUT AND TRUNCATED SYSTEMS

Consider continuous time Markov chain $X(t) = (C(t); N(t))$, $C(t) \in \{0, 1, \ldots, c\}$, $N(t) \in \{0, 1, \ldots\}$, which is defined by its infinitesimal characteristics $a_{(i,j)(i',j')}$, (i, j), $(i', j') \in S(X) = \{0, 1, \ldots, c\} \times \{0, 1, \ldots\}$:

1) if $i = \{0, 1, \ldots, c-1\}$, then

$$a_{(i,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (i+1, j); \\ j\mu, & (i',j') = (i+1, j-1); \\ i\nu_j, & (i',j') = (i-1, j); \\ -[\lambda + j\mu + i\nu_j], & (i',j') = (i, j); \\ 0, & \text{otherwise.} \end{cases}$$

2) if $i = c$, then

$$a_{(c,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (c, j+1); \\ c\nu_j, & (i',j') = (c-1, j); \\ -[\lambda + c\nu_j], & (i',j') = (c, j); \\ 0, & \text{otherwise.} \end{cases}$$

Process $X(t)$ describes the behavior of the following system. The incoming flow of events is Poisson with rate λ. There are c identical servers. If there is any free server the request is served immediately. Service time is exponentially distributed random variable with parameter ν_j that depends on the current queue of repeated calls length. If request finds all servers busy it tries to get service in a random period of time that has exponential distribution with parameter μ. The number of busy servers at any time t is defined by the first component of $X(t)$ and the number of retrials Ц by the second.

Let us find up the condition of $X(t), t > 0$ stationary mode existence.

Lemma 1. *Let $\nu = \varliminf_{j\to\infty} \nu_j$. Then if $\frac{\lambda}{c\nu} < 1$ process $X(t)$ – is ergodic and its boundary distribution $\pi_{ij}, (i,j) \in S(X)$ coincides with a single stationary.*

Proof. Consider the following Lyapunov functions $\varphi(i,j) = \alpha i + j, (i,j) \in S(X)$, where parameter α will be determined later on. Then the mean drifts

$$y_{ij} = \sum_{(i',j') \neq (i,j)} a_{(i,j)(i',j')}(\varphi(i',j') - \varphi(i,j)$$

are given by

$$y_{ij} = \begin{cases} \lambda\alpha - i\nu_j\alpha + j\mu(\alpha - 1), & 0 \le i \le c-1, \\ \lambda - c\nu_j\alpha, & i = c. \end{cases}$$

When $\frac{\lambda}{c\nu} < 1$ for any value of $\alpha \in (\frac{\lambda}{c\nu}, 1)$ there exists such $\varepsilon > 0$ that $y_{ij} < -\varepsilon$ for all $(i,j) \in S(X)$ excluding finite number of states (i,j). So the assumptions of Tweedie's theorem[[?], p.97] are hold for functions $\varphi(i,j) = \alpha i + j, \alpha \in (\frac{\lambda}{c\nu}, 1)$. □

Consider truncated system. It functions in the same way as an input system but has the limitation on maximum number of retrials. This means that incoming calls are lost when all servers are busy and there are M sources of repeated calls already. Formally the system is described by the Markov chain $X(t, M) = (C(t, M); N(t, M))$,

where $C(t,M) \in \{0,1,\dots,c\}$, $N(t,M) \in \{0,1,\dots,M\}$ with infinitesimal characteristics $a^{(M)}_{(i,j)(i',j')}, (i,j),(i',j') \in S(X,M) = \{0,1,\dots,c\} \times \{0,1,\dots,M\}$:

1) if $i = \{0,1,\dots,c-1\}, j = \{0,1,\dots,M\}$, then

$$a^{(M)}_{(i,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (i+1,j); \\ j\mu, & (i',j') = (i+1,j-1); \\ i\nu_j, & (i',j') = (i-1,j); \\ -[\lambda + j\mu + i\nu_j, & (i',j') = (i,j); \\ 0, & \text{otherwise.} \end{cases}$$

2) if $i = c, j = \{0,1,\dots,M-1\}$, then

$$a^{(M)}_{(c,j)(i',j')} = \begin{cases} \lambda, & (i',j') = (c,j+1); \\ c\nu_j, & (i',j') = (c-1,j); \\ -[\lambda + c\nu_j], & (i',j') = (c,j); \\ 0, & \text{otherwise.} \end{cases}$$

3) if $i = c, j = M$, then

$$a^{(M)}_{(c,M)(i',j')} = \begin{cases} c\nu_M, & (i',j') = (c-1,M); \\ -c\nu_M, & (i',j') = (c,M); \\ 0, & \text{otherwise.} \end{cases}$$

The state space $S(X,M)$ of process $X(t,M)$ is finite thus the stationary mode always exists and by $\pi_{ij}(M), (i,j) \in S(X,M)$ we define its stationary probabilities.

Let us consider the service process of the truncated system in more detail.

3. STATIONARY PROBABILITIES OF THE TRUNCATED SYSTEM

For the given system stationary probabilities satisfy the following set of Kolmogorov equations and normalizing condition:

$$[\lambda + j\mu + i\nu_j]\pi_{ij}(M) = (j+1)\mu\pi_{i-1j+1}(M) + \lambda\pi_{i-1j}(M) + (i+1)\nu_j\pi_{i+1j}(M), \quad (1)$$

$$j = 0,\dots,M-1, i = 0,...,c-1,$$

$$[\lambda + M\mu + i\nu_M]\pi_{iM}(M) = \lambda\pi_{i-1M}(M) + (i+1)\nu_M\pi_{i+1M}(M), i = 0,...,c-1, \quad (2)$$

$$[\lambda + c\nu_j]\pi_{cj}(M) = (j+1)\mu\pi_{c-1j+1}(M) + \lambda\pi_{c-1j}(M) + \lambda\pi_{cj-1}(M), j = 0,\dots,M-1,$$

$$c\nu_M\pi_{cM}(M) = \lambda\pi_{c-1M}(M) + \lambda\pi_{cM-1}(M), \quad (3)$$

$$\sum_{i=0}^{c}\sum_{j=0}^{M}\pi_{ij}(M) = 1.$$

Consider the following definitions:

$$e_i(n) = (\delta_{i0},\ \delta_{i1},\ \ldots,\ \delta_{in-1})^T, \quad \delta_{ij} = \begin{cases} 1, i = j; \\ 0, i \neq j; \end{cases}$$

$1(c)$ - vector of length that consists from 1,

$$A_j = \|a^j_{ik}\|^{c-1}_{i,k=0}, \quad a^j_{ik} = \begin{cases} -\lambda, & k = i-1, \\ \lambda + j\mu + i\nu_j, & k = i, \\ -(i+1)\nu_j, & k = i+1, \\ 0, & \text{otherwise}; \end{cases}$$

if $i \neq 0, c-1$. In case $i = 0$

$$a^j_{0k} = \begin{cases} \lambda + j\mu, & k = 0, \\ -\nu_j, & k = 1, \\ 0, & \text{otherwise}; \end{cases}$$

and if $i = c-1$

$$a^j_{c-1k} = \begin{cases} -\lambda, & k = c-2, \\ \lambda + j\mu + (c-1)\nu_j, & k = c-1, \\ 0, & \text{otherwise}; \end{cases}$$

$$B_j = \|b^j_{ik}\|^{c-1}_{i,k=0}, \quad b^j_{ik} = \begin{cases} (j+1)\mu, & k = i-1, \\ 0, & \text{otherwise}; \end{cases}$$

if $i \neq 0, c-1$. In case $i = 0, b^j_{0k} = 0, k = 0, 1, \ldots, c-1$, and if $i = c-1$

$$b^j_{c-1k} = \begin{cases} \frac{c(j+1)\mu\nu_j}{\lambda}, & k \neq c-2, \\ \frac{(j+1)\mu[\lambda + c\nu_j]}{\lambda}, & k = c-2; \end{cases}$$

$$C = \|c_{ik}\|^{c-1}_{i,k=0}, \quad c_{ik} = \begin{cases} 1, & k = 0, i = 0, \\ a^M_{i-1k}, & \text{otherwise}; \end{cases}$$

$$\Phi_j = \left(\prod_{i=j}^{M-1} A_i^{-1} B_i \right) C^{-1} e_0(c).$$

Vector Φ_j is defined correctly by the last equation. As $|C| = (-1)^{c-1}(c-1)\nu_M^{c-1} \neq 0$ so C^{-1} always exists. Matrices $A_j, j = 0, 1, \ldots, M$ are not singular because they satisfy the Adamar column condition ([3] p. 406).

Probabilities $\pi_{ij}(M), (i,j) \in S(X, M)$ can be explicitly expressed in terms of the system'Ÿs parameters.

Theorem 1. *Stationary probabilities of the system are defined by the following set of equations:*

$$\pi_j(M) = \Phi_j \pi_{0M}(M), j = 0, \ldots, M,$$

$$\pi_{cj}(M) = \frac{(j+1)\mu}{\lambda} 1(c)^T \Phi_{j+1} \pi_{0M}(M), j = 0, \ldots, M-1,$$

$$\pi_{cM}(M) = \frac{\lambda e_{c-1}^T(c) + M\mu 1(c)^T}{c\nu_M} C^{-1} e_0(c) \pi_{0M}(M),$$

where $\pi_j(M) = (\pi_{0j}(M),\ \pi_{1j}(M),\ \ldots,\ \pi_{c-1j}(M))^T$,

$$\pi_{0M}(M) = \left\{ \sum_{j=0}^{M} \left(1 + \frac{j\mu}{\lambda}\right) 1(c)^T \Phi_j + \frac{\lambda e_{c-1}^T(c) + M\mu 1(c)^T}{c\nu_M} \Phi_M \right\}^{-1}.$$

It should be noticed that in case of $c = 1, 2$ the above formulas turn into equations of the scalar type [4].

Next we will show that the system with truncated state space approximates the input system.

4. APPROXIMATION RESEARCH

To proof that characteristics of the finite system approximate characteristics of the input system let us use stochastic order concept [5].

Lemma 2. *If conditions of lemma 1 are true and :*

1) if $X(0, M) \leq_{st} X(0)$, *then* $X(t, M) \leq_{st} X(t)$ *for all* $t \geq 0$ *and* $X(M) \leq_{st} X$, *where* $X = (C, N)$, $X(M) = (C(M), N(M))$ *– random vectors distributed as* $\pi_{ij}, (i, j) \in S(X)$ *and* $\pi_{ij}(M), (i, j) \in S(X, M)$ *correspondingly;*

2) if $X(0, M) \leq_{st} X(0, M+1)$, *then* $X(t, M) \leq_{st} X(t, M+1)$ *for all* $t \geq 0$ *and* $X(M) \leq_{st} X(M+1)$.

Lemma 2 appears from the results of stochastic order of migration processes. It leades to the following theorem.

Theorem 2. *If conditions of lemma 1 take place, then for any* $(i, j) \in S(X)$ $\pi_{ij} = \lim_{M \to \infty} \pi_{ij}(M)$.

We have already shown that the probabilities $\pi_{ij}(M)$ tend to π_{ij} as the value of M increases. It is interesting to find the value of difference between them. Let us examine it for some pretty general types of policies.

Theorem 3. *If conditions of lemma 1 take place and the sequence* ν_j *is strictly monotone, then for* $(i, j) \in \{1, \ldots, c\} \times \{0, 1, \ldots\}$

$$0 \leq \bar{\pi}_{ij} - \bar{\pi}_{ij}(M) \leq \lambda \left(\nu_j - \nu_{j-1}\right)^{-1} \frac{\frac{1}{M!}\left(\frac{\lambda}{c\mu}\right)^M \prod_{\alpha=0}^{M} \frac{\lambda + \alpha\mu}{\nu_\alpha}}{\sum_{\beta=0}^{M} \frac{1}{\beta!}\left(\frac{\lambda}{c\mu}\right)^\beta \left[c + \frac{\lambda + \beta\mu}{\nu_\beta}\right] \prod_{\alpha=0}^{\beta-1} \frac{\lambda + \alpha\mu}{\nu_\alpha}},$$

where $\bar{\pi}_{ij} = \sum\limits_{\alpha \geq i, \beta \geq j} \pi_{\alpha\beta} = P(C \geq i, N \geq j)$, $\bar{\pi}_{ij}(M) = \sum\limits_{\alpha \geq i, \beta \geq j} \pi_{\alpha\beta}(M) = P(C(M) \geq i, N(M) \geq j)$.

Theorem 4. *If conditions of lemma 1 take place and* $\nu_j \leq \nu_{j+1}, j = 0, 1, \ldots,$ $\lambda < c\nu_0$ *then*

$$\sup_{(i,j) \in S(X)} (\bar{\pi}_{ij} - \bar{\pi}_{ij}(M)) \leq \frac{\frac{c[(\lambda + c\mu)\lambda + (M+c)\nu_0]}{M!\nu_0} (\frac{\lambda}{c\mu})^{M+1} \prod\limits_{\alpha=0}^{M} \frac{\lambda + \alpha\mu}{\nu_\alpha}}{(c\nu_0 - \lambda) \sum\limits_{\beta=0}^{M} \frac{1}{\beta!} (\frac{\lambda}{c\mu})^{\beta} [c + \frac{\lambda + \beta\mu}{\nu_\beta}] \prod\limits_{\alpha=0}^{\beta-1} \frac{\lambda + \alpha\mu}{\nu_\alpha}},$$

In monograph of Failn and Tempelton the differences between some major integral functionals of the input and the truncated system were estimated in case of uncontrolled retrial queues [5]. In case of controlled retrial queues similar results can be found but such an estimate strongly relates on the control policy properties. We have provided such estimates for few of them.

REFERENCES

1. *Artalejo J. R., Gomez-Corral A.* Retrial queueing systems // Springer-Verlag, Berlin, 2008.
2. *Semenova O. V., Dudin A. N.* M/M/N queueing system with controlled service mode and disaster // Avtomatika i Vychislitel'naya Tekhnika. 2007. No. 6. P. 72–80.
3. *Gantmaher F. R.* Matrix theory //Nauka, Moscow, 1967.
4. *Lebedev E. O., Ponomarov V. D.* Optimization of finitesource retrial queues // Bulletin of Kiev University. 2008. V. 2. P. 91–97.
5. *Falin G. I., Templeton J. G. C.* Retrial Queues // Chapman and Hall, London, 1997.

INVESTIGATION OF QUEUING SYSTEM $GI^{(2)}|M_2|\infty$

I. Sinyakova[1], S. Moiseeva[2]

[1] *The branch of Kemerovo state university in Anzhero-Sudzhensk,*
[2] *Tomsk state university*
[1] *Anzhero-Sudzhensk, Russia*
[2] *Tomsk, Russia*
[1] Irinka_asf@mail.ru

We offer the model of applications parallel service in queuing system (QS) which consists of two units of service with an unlimited number of servers. A recurrent stream of binary applications enters to an input to the system. We recorded the approximate (asymptotic) equality of the first and second orders for the respective blocks of the system.

The research is made under supporting of ADAP "Development of the higher education scientific potential" (2009-2010 years) of the Federal education agency of RF concerning with the project "Investigation methods formulation of non-Markov queues systems and their appliance in complex economic systems and computer transmission network".

Keywords: parallel service, systems with an unlimited number of servers, a recurrent stream, the method of asymptotic analysis.

1. INTRODUCTION

Currently, attention to the queuing theory is largely stimulated by the need to apply the results of this theory to important practical problems arising in connection with the rapid development of communication systems, the emergence of information and computing systems, appearance and complexity of various technological systems, the creation of automated control systems.

At the present stage of development of queuing theory, one of the urgent tendencies is the queuing systems investigation with batch arrival of applications and parallel service [1, 2]. Range of application of the this QS is quite extensive, for example, modeling of modern information and computer systems requires take into account the packet traffic, as well as one of the basic principles for the design of modern computer networks - parallel information processing [3, 4]. Therefore there is a need to develop new mathematical models of queuing systems, such as systems with extraordinary incoming flow and the various service options, including two or more units of service.

2. PROBLEM STATEMENT

We consider a QS $GI^{(2)}|M_2|\infty$ with two service blocks, each of which contains an unlimited number of servers.

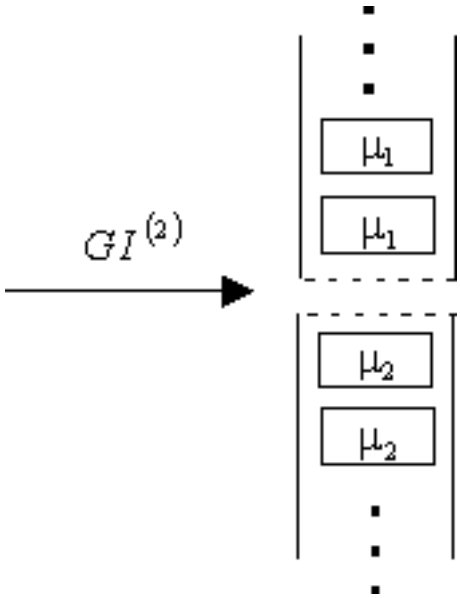

Fig. 1. QS with parallel services of multiple applications $GI^{(2)}|M_2|\infty$

A recurrent flow [5] of dual applications specified distribution function A(x) - length of intervals between the moments of events in the concerned stream enters to the input of the system. Service intervals of various applications are stochastically independent and identically distributed in each block and have an exponential distribution with μ_1 and $\mu_2$parameters, respectively. Received applications takes any of the free servers and after completion of the service the application leaves the system.

3. MATHEMATICAL MODEL

We define $i_k(t)$ - number of requests in the k-th block of service. Because the incoming flow is non-Poisson, the two-dimensional process $\{i_1(t),i_2(t)\}$ is non-Markov, and therefore we consider an additional component of z(t), equal to the length of the interval from t to themoment of the next event in the recurrent input stream.

For the concerned system a three-dimensional random process $\{z(t),i_1(t),i_2(t)\}$ is Markov, so its probability distribution

$$P\left(z,i_1,i_2,t\right)=P\{z\left(t\right)<z,i_1\left(t\right)=i_1,i_2\left(t\right)=i_2\}$$

We can write equality

$$P(z-\Delta t,i_1,i_2,t+\Delta t)=\{P(z,i_1,i_2,t)-P(\Delta t,i_1,i_2,t)\}\left(1-i_1\mu_1\Delta t\right)\left(1-i_2\mu_2\Delta t\right)+$$

$$+P(\Delta t,i_1-1,i_2-1,t)A(z)+P(z,i_1,i_2+1,t)\left(i_2+1\right)\mu_2\Delta t+$$

$$+P(z,i_1+1,i_2,t)\,(i_1+1)\,\mu_1\Delta t+o(\Delta t),$$

from which we can readily obtain a system of differential equations of Kolmogorov [6]

$$\frac{\partial P(z,i_1,i_2,t)}{\partial t}=\frac{\partial P(z,i_1,i_2,t)}{\partial z}-\frac{\partial P(0,i_1,i_2,t)}{\partial z}-P(z,i_1,i_2,t)i_1\mu_1-P(z,i_1,i_2,t)i_2\mu_2+$$

$$+P(z,i_1+1,i_2,t)\,(i_1+1)\,\mu_1+P(z,i_1,i_2+1,t)\,(i_2+1)\,\mu_2+\frac{\partial P(0,i_1-1,i_2-1,t)}{\partial z}A(z).$$

For the stationary probability distribution this system can be rewritten as

$$\frac{\partial P(z,i_1,i_2)}{\partial z}-\frac{\partial P(0,i_1,i_2)}{\partial z}-P(z,i_1,i_2)i_1\mu_1-P(z,i_1,i_2)i_2\mu_2+$$

$$+P(z,i_1+1,i_2)\,(i_1+1)\,\mu_1+P(z,i_1,i_2+1)\,(i_2+1)\,\mu_2+\frac{\partial P(0,i_1-1,i_2-1)}{\partial z}A(z)=0.$$

Defining $H(z,u_1,u_2)=\sum\limits_{i_1=0}^{\infty}\sum\limits_{i_2=0}^{\infty}e^{ju_1i_1}e^{ju_2i_2}P(z,i_1,i_2)$, where $j=\sqrt{-1}$ imaginary unit and taking into account that

$$\frac{\partial H(z,u_1,u_2)}{\partial u_1}=j\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}i_1e^{ju_1i_1}e^{ju_2i_2}P(z,i_1,i_2),$$

$$\frac{\partial H(z,u_1,u_2)}{\partial u_2}=j\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}i_2e^{ju_1i_1}e^{ju_2i_2}P(z,i_1,i_2),$$

for the functions H(z,u_1,u_2) we obtain the basic equation for investigating the system $GI^{(2)}|M_2|\infty$

$$\frac{\partial H\,(z,u_1,u_2)}{\partial z}+\frac{\partial H\,(0,u_1,u_2)}{\partial z}\left\{e^{j(u_1+u_2)}A\,(z)-1\right\}-j\mu_1\left(e^{-ju_1}-1\right)\frac{\partial H\,(z,u_1,u_2)}{\partial u_1}-$$

$$-\,j\mu_2\left(e^{-ju_2}-1\right)\frac{\partial H\,(z,u_1,u_2)}{\partial u_2}=0. \qquad (1)$$

We define additional conditions as

$$H(z,0,0)=R(z),$$

where R(z) – stationary distribution of the z(t). Solution H(z,u_1,u_2) defines the characteristic function of the number of servers employed in the stationary state in each block of the system $GI^{(2)}|M_2|\infty$ as equalities

$$Me^{ju_1i_1(t)}=H(\infty,u_1,0),$$

$$Me^{ju_2i_2(t)}=H(\infty,0,u_2).$$

4. THE METHOD OF ASYMPTOTIC ANALYSIS

4.1. Asymptotic form of the first order. For a complete analysis of the studied queuing systems apply the method of asymptotic analysis [5].

We consider the basic equation for the characteristic function (1):

$$\frac{\partial H\left(z,u_1,u_2\right)}{\partial z}+\frac{\partial H\left(0,u_1,u_2\right)}{\partial z}\left(e^{j(u_1+u_2)}A\left(z\right)-1\right)-j\mu_1\left(e^{-ju_1}-1\right)\frac{\partial H\left(z,u_1,u_2\right)}{\partial u_1}-$$

$$-j\mu_2\left(e^{-ju_2}-1\right)\frac{\partial H\left(z,u_1,u_2\right)}{\partial u_2}=0,$$

which will be solved in the asymptotic conditions of the growing service-time, believing that μ_1,μ_2 →0. We define $\mu_1=\varepsilon$, $\mu_2=\varepsilon$q and do changes in equation (1)

$$u_1=\varepsilon w_1,\quad u_2=\varepsilon w_2,\quad H(z,u_1,u_2)=F_1(z,w_1,w_2,\varepsilon), \tag{2}$$

In the result we get the equation for F_1(z,w_1,w_2,ε) in the following form:

$$\frac{\partial F_1(z,w_1,w_2,\varepsilon)}{\partial z}+\frac{\partial F_1(0,w_1,w_2,\varepsilon)}{\partial z}\left(e^{j(w_1+w_2)}A\left(z\right)-1\right)-$$

$$-\,j\left(e^{-j\varepsilon w_1}-1\right)\frac{\partial F_1(z,w_1,w_2,\varepsilon)}{\partial w_1}-j\left(e^{-j\varepsilon w_2}-1\right)\frac{\partial F_1(z,w_1,w_2,\varepsilon)}{\partial w_2}=0. \tag{3}$$

We prove the following statement:

Theorem 1. *The limit (when ε →0) solution of equation* (2) *has the form:*

$$F_1(z,w_1,w_2)=R(z)\mathbf{e}^{\{j(w_1+w_2)\lambda\}},$$

where R(z) - stationary probability distribution of random process values{z(t)}, and the parameter λ is given by: $\frac{\partial R(0)}{\partial z}=\lambda$

Proof. The proof 1. In the equation (3) limiting transition is executed at $\varepsilon\to 0$, we will receive that F_1(z,w_1,w_2) the equation decision is

$$\frac{\partial F_1(z,w_1,w_2)}{\partial z}+\frac{\partial F_1(0,w_1,w_2)}{\partial z}\{A(z)-1\}=0,$$

which defines a vector function R(z), therefore

$$F_1(z,w_1,w_2)=R(z)\Phi_1\left(w_1,w_2\right) \tag{4}$$

We find the scalar function Φ_1(w_1,w_2) as follows. In the equation (3) limiting transition is performed at z → ∞, we will receive an equality

$$\frac{\partial F_1(0,w_1,w_2,\varepsilon)}{\partial z}\left(e^{j(w_1+w_2)\varepsilon}-1\right)-j\left(e^{-j\varepsilon w_1}-1\right)\frac{\partial F_1(\infty,w_1,w_2,\varepsilon)}{\partial w_1}-$$

$$-j\left(e^{-j\varepsilon w_2}-1\right)\frac{\partial F_1(\infty,w_1,w_2,\varepsilon)}{\partial w_2}=0,$$

dividing the left and right parts of the equation by ε and executing limiting transition at $\varepsilon \to 0$, we will receive the equality

$$\frac{\partial F_1(0,w_1,w_2)}{\partial z} j\left(w_1+w_2\right)+j^2 w_1 \frac{\partial F_1(\infty,w_1,w_2)}{\partial w_1}+j^2 w_2 \frac{\partial F_1(\infty,w_1,w_2)}{\partial w_2}=0,$$

in which we substitute the expression (4) and write down the differential equation in partial derivatives

$$\frac{\partial \Phi_1\left(w_1,w_2\right)}{\partial w_1}+j^2 w_2 \frac{\partial \Phi_1\left(w_1,w_2\right)}{\partial w_2}=j\left(w_1+w_2\right) \frac{\partial R(0)}{\partial z} \Phi_1\left(w_1,w_2\right)=$$

$$=j\left(w_1+w_2\right) \lambda \Phi_1\left(w_1,w_2\right),$$

whose solution is satisfying the initial condition $\Phi_1(0,0)=1$ and it has the form

$$\Phi_1\left(w_1,w_2\right)=\mathbf{e}^{\{j\lambda(w_1+w_2)\}}.$$

We get $F_1(z,w_1,w_2)=R(z)\mathbf{e}^{\{j\lambda(w_1+w_2)\}}$.

The theorem is proved. □

Using (2) and (4) we can write the approximate (asymptotic) equality

$$H\left(z,u_1,u_2\right)=R(z)\mathbf{e}^{\left\{j\lambda\left(\frac{u_1}{\varepsilon}+\frac{u_2}{\varepsilon}\right)\right\}}=R(z)\mathbf{e}^{\left\{j\lambda\left(\frac{u_1}{\mu_1}+\frac{u_2}{\mu_2 q}\right)\right\}}$$

from which we can obtain following expression for the characteristic function of process $\{i_1(t),i_2(t)\}$ in the stationary state,

$$Me^{ju_1 i_1(t)}=H\left(\infty,u_1,0\right)=\mathbf{e}^{\left\{ju_1\frac{\lambda}{\mu_1}\right\}}, \quad Me^{ju_2 i_2(t)}=H\left(\infty,0,u_2\right)=\mathbf{e}^{\left\{ju_2\frac{\lambda}{\mu_2 q}\right\}}.$$

The received equations is called the first order asymptotic form for the service blocks of systems service $GI^{(2)}|M_2|\infty$.

For a more detailed investigation, we consider the second order asymptotic form.

4.2. Asymptotic form of the second order. In equation (1) we make the substitution

$$H\left(z,u_1,u_2\right)=H_2(z,u_1,u_2)\mathbf{e}^{\left\{j\lambda\left(\frac{u_1}{\mu_1}+\frac{u_2}{\mu_2 q}\right)\right\}},$$

which later becomes

$$\frac{\partial H_2(z,u_1,u_2)}{\partial z}+\frac{\partial H_2(0,u_1,u_2)}{\partial z}\left(\mathbf{e}^{\{j(u_1+u_2)\}}A\left(z\right)-1\right)+$$

$$+j\mu_1\left(1-e^{-ju_1}\right)\frac{\partial H_2(z,u_1,u_2)}{\partial u_1}+j\mu_2 q\left(1-e^{-ju_2}\right)\frac{\partial H_2(z,u_1,u_2)}{\partial u_2}-$$

$$-\lambda\left(1-e^{-ju_1}\right)H_2(z,u_1,u_2)-\lambda\left(1-e^{-ju_2}\right)H_2(z,u_1,u_2)=0.$$

We define $\mu_1=\varepsilon^2$, $\mu_2=\varepsilon^2 q$ in the last equation and make substitutions

$$u_1=\varepsilon w_1, \quad u_2=\varepsilon w_2, \quad H_2(z,u_1,u_2)=F_2(z,w_1,w_2,\varepsilon),$$

we get the result

$$\frac{\partial F_2(z,w_1,w_2,\varepsilon)}{\partial z}+\frac{\partial F_2(0,w_1,w_2,\varepsilon)}{\partial z}\left(e^{j\varepsilon(w_1+w_2)}A\left(z\right)-1\right)+$$
$$+j\varepsilon\left(1-e^{-j\varepsilon w_1}\right)\frac{\partial F_2(z,w_1,w_2,\varepsilon)}{\partial w_1}+j\varepsilon\left(1-e^{-j\varepsilon w_2}\right)\frac{\partial F_2(z,w_1,w_2,\varepsilon)}{\partial w_2}-$$
$$-\lambda\left(1-e^{-j\varepsilon w_1}\right)\partial F_2(z,w_1,w_2,\varepsilon)-\lambda\left(1-e^{-j\varepsilon w_2}\right)\partial F_2(z,w_1,w_2,\varepsilon)=0. \qquad (5)$$

The following theorem is proved similarly theorem for first order asymptotic form.

Theorem 2. *The limit (when $\varepsilon\to 0$) solution of equation* (5) *has the form*

$$F_2(z,w_1,w_2)=R(z)\mathbf{e}^{\left\{\frac{(j(w_1+w_2))^2}{2}\kappa_2\right\}},$$

where R(z) - stationary probability distribution of the random process values {z(t)}, the parameter λ is given by: $\frac{\partial R(0)}{\partial z}=\lambda$, the value $\kappa_2$is determined as $\kappa_2=\lambda+\frac{\partial f_2(0)}{\partial z}$.

The vector function $f_2(z)$ satisfies the condition $f_2(\infty)=0$ and it is a solution of the equation

$$\frac{\partial f_2(z)}{\partial z}+\frac{\partial f_2(0)}{\partial z}\left(A(z)-1\right)+\frac{\partial R(0)}{\partial z}A(z)-\lambda R(z)=0.$$

Also the asymptotic (approximate) equality can be written

$$H_2(z,u_1,u_2)=F_2(z,w_1,w_2\varepsilon)\approx F_2(z,w_1,w_2)=$$
$$=R(z)\mathbf{e}^{\left\{j^2\left[\kappa_2\frac{(w_1+w_2)^2}{2}-\frac{\lambda w_1 w_2}{2}\right]\right\}}=$$
$$=R(z)\mathbf{e}^{\left\{\frac{j^2\kappa_2}{2}\left(\frac{u_1}{\sqrt{\mu_1}}+\frac{u_2}{\sqrt{\mu_2 q}}\right)^2-j^2\frac{\lambda u_1 u_2}{2\sqrt{\mu_1\mu_2 q}}\right\}},$$

from which we can obtain following expression for the characteristic function H(z,u_1,u_2)

$$H(z,u_1,u_2)=R(z)\mathbf{e}^{\left\{j\lambda\left(\frac{u_1}{\mu_1}+\frac{u_2}{\mu_2 q}\right)+\frac{j^2\kappa_2}{2}\left(\frac{u_1}{\sqrt{\mu_1}}+\frac{u_2}{\sqrt{\mu_2 q}}\right)^2-j^2\frac{\lambda u_1 u_2}{2\sqrt{\mu_1\mu_2 q}}\right\}},$$

Then the marginal characteristic functions have the form

$$Me^{ju_1 i_1(t)}=H(\infty,u_1,0)=\mathbf{e}^{\left\{j\lambda\frac{u_1}{\mu_1}+\frac{(ju_1)^2}{2\mu_1}\kappa_2\right\}},$$
$$Me^{ju_2 i_2(t)}=H(\infty,0,u_2)E=\mathbf{e}^{\left\{j\lambda\frac{u_2}{\mu_2 q}+\frac{(ju_2)^2}{2\mu_2 q}\kappa_2\right\}}.$$

The received equations is called second order asymptotic form for service blocks of the system $GI^{(2)}|M_2|\infty$.

5. CONCLUSION

So we have constructed a model of multiple parallel service requests of recurrent flow. We investigate two-dimensional stochastic process characterizing the number of engaged servers in the first and the second service block. We have found expressions for the characteristic functions at the asymptotic conditions of the growing claims service-time.

REFERENCES

1. *Tatashev A. G.* Queuing system group with group arrival and inverse discipline // Cybernetics and of systems analyses 1995. № 6. P. 163–165.
2. *Chechelnitsky A. A., Kucherenko O. B.* Stationary characteristic of queuing system with parallel service // Science journal // Minsk, 2009. P. 262–268.
3. *Endrus G. R.* Basics of multithreaded, parallel and distributed software technology: Transl. from English. //M.: Publishing house "Williams" 2003. P. 512.
4. *Toporkov V. V.* Models of grid computing. // M.: FIZMATLIT, 2004. P. 320.
5. *Nazarov A. A., Moiseeva S. P.* Method of asymptotic analysis in theory of queuing systems. // Tomsk: NTL, 2006. P. 112.
6. *Kleinrock L.* Queuing systems. Transl. from English Glushko. // M. : Mechanical engeneering, 1979. P. 432.

О ТОЧНОСТИ ОДНОГО МЕТОДА НАХОЖДЕНИЯ ОЖИДАЕМЫХ ДОХОДОВ В НМ-СЕТЯХ

С. Статкевич, М. Маталыцкий

Гродненский государственный университет им. Я. Купалы
Гродно, Беларусь
sstat@grsu.by

В работе рассматривается экспоненциальная НМ-сеть с ограниченными временами ожидания заявок в очередях систем обслуживания и случайными доходами от переходов между ее состояниями. Обсуждается использование приближенной формулы $M\min(k_i(t), m_i) \approx \min(M\{k_i(t)\}, m_i)$, где $k_i(t)$ и m_i - соответственно число заявок и число линий обслуживания в $i-$ой системе, $i = \overline{1,n}$, при нахождении ожидаемых доходов.

Ключевые слова: НМ-сеть, ожидаемые доходы, неравенство Иенсена.

1. ВВЕДЕНИЕ

Рассмотрим открытую сеть массового обслуживания (МО) с однотипными заявками, состоящую из n СМО $S_1, S_2, \ldots, S_n$. Под состоянием сети понимается вектор $k(t) = (k_1(t), k_2(t), \ldots, k_n(t))$, где $k_i(t)-$число заявок в системе S_i в момент времени t, $i = \overline{1,n}$. В сеть поступает простейший поток заявок с интенсивностью λ. Интенсивность обслуживания заявок в момент времени t $\mu_i(k_i(t))$ в системе S_i зависит от числа заявок в ней (в очереди и на обслуживании) $k_i(t)$, $i = \overline{1,n}$. Пусть $p_{0j}-$вероятность поступления заявок в систему S_j из внешней среды, $\sum\limits_{j=1}^{n} p_{0j} = 1$; $p_{ij}-$вероятность перехода заявки после обслуживания в системе S_i в систему S_j, $p_{00} = 0$, $\sum\limits_{j=0}^{n} p_{ij} = 1$, $i = \overline{1,n}$. Заявка при переходе из одной СМО в другую приносит последней системе некоторый случайный доход и соответственно доход первой системы уменьшается на эту СВ.

Длительность пребывания заявок в очереди $i-$ой СМО является СВ, распределенной по экспоненциальному закону с параметром $\theta_i(k_i(t))$, и не зависит от других факторов, например, от времени пребывания в очереди других заявок. Заявка, время ожидания которой в очереди S_i истекло, переходит в систему S_j с вероятностью q_{ij}, $q_{ii} = 0$, $i = \overline{1,n}$, $j = \overline{0,n}$. Матрицы $P = \|p_{ij}\|_{(n+1)\times n}$ и $Q = \|q_{ij}\|_{n\times(n+1)}$ являются матрицами переходов неприводимых марковских цепей.

2. НАХОЖДЕНИЕ ОЖИДАЕМЫХ ДОХОДОВ

Обозначим через $V_i(t)$ доход системы S_i в момент времени t и пусть в начальный момент времени ее доход равен $V_i(0) = v_{i0}$. Доход этой же СМО в момент времени $t + \Delta t$ можно представить в виде

$$V_i(t + \Delta t) = V_i(t) + \Delta V_i(t, \Delta t), \tag{1}$$

где $\Delta V_i(t, \Delta t)$ - изменение дохода системы S_i на интервале времени $[t, t + \Delta t)$. Для величины $\Delta V_i(t, \Delta t)$ справедливо соотношение [1]:

$$\Delta V_i(t, \Delta t) = \begin{cases} r_{i0} + r_i\Delta t & \text{с вероятностью} \quad \lambda p_{0i}\Delta t + o(\Delta t), \\ -R_{i0} + r_i\Delta t & \text{с вероятностью} \quad \mu_i(k_i(t))u(k_i(t))p_{i0}\Delta t + o(\Delta t), \\ r_{ji} + r_i\Delta t & \text{с вероятностью} \quad \mu_j(k_j(t))u(k_j(t))p_{ji}\Delta t + o(\Delta t), \\ -R_{ij} + r_i\Delta t & \text{с вероятностью} \quad \mu_i(k_i(t))u(k_i(t))p_{ij}\Delta t + o(\Delta t), \\ -H_{i0} + r_i\Delta t & \text{с вероятностью} \quad \theta_i(k_i(t))u(k_i(t))q_{i0}\Delta t + o(\Delta t), \\ h_{ji} + r_i\Delta t & \text{с вероятностью} \quad \theta_j(k_j(t))u(k_j(t))q_{ji}\Delta t + o(\Delta t), \\ -H_{ij} + r_i\Delta t & \text{с вероятностью} \quad \theta_i(k_i(t))u(k_i(t))q_{ij}\Delta t + o(\Delta t), \\ r_i\Delta t & \text{с вероятностью} \quad 1 - \Big[\lambda p_{0i} + \big(\mu_i(k_i(t)) + \\ & \qquad + \theta_i(k_i(t))\big)u(k_i(t)) + \sum\limits_{\substack{j=1 \\ (j\neq i)}}^{n} \big(\mu_j(k_j(t))p_{ji} + \\ & \qquad + \theta_j(k_j(t))q_{ji}\big)u(k_j(t))\Big]\Delta t + o(\Delta t), \end{cases} \tag{2}$$

где r_{0i} - доход системы S_i, который приносит ей заявка, поступающая из внешней среды, $M\{r_{0i}\} = a_{0i}$, $-R_{i0}$ - убытки системы S_i, приносящие ей заявкой, уходящей из нее во внешнюю среду, $M\{R_{i0}\} = b_{i0}$, $r_i\Delta t$ - увеличение дохода системы S_i за время Δt, когда состояние сети не меняется, $-H_{i0}$ - убытки системы S_i, приносящие ей заявкой которая уходит из нее во внешнюю среду, не дождавшись обслуживания, $i = \overline{1, n}$; r_{ji} - доход системы S_i, который приносит ей заявка, поступающая из системы S_j, $M\{r_{ji}\} = a_{ji}$, $-R_{ij}$ - убытки системы S_i, приносящие ей заявкой, уходящей из нее в систему S_j, $M\{R_{ij}\} = b_{ij}$, h_{ji} - доход системы S_i, приносящий ей заявкой, которая поступает из системы S_j, не дождавшись в ней обслуживания, $M\{h_{ji}\} = \overline{h}_{ji}$, $-H_{ij}$ - убытки системы S_i, приносящие ей заявкой, которая, не дождавшись в ней обслуживания, переходит в систему S_j, $M\{H_{ij}\} = \overline{H}_{ij}$, $i \neq j$, $i, j = \overline{1, n}$.

Записывая, учитывая (2), выражения для $M\{\Delta V(t, \Delta t)/k(t)\}$ при фиксированной реализации процесса $k(t)$, и усредняя по $k(t)$ с учетом условия нормировки $\sum\limits_k P(k(t) = k) = 1$, для изменения ожидаемого дохода системы S_i получаем

$$M\{\Delta V_i(t, \Delta t)\} = \sum_k P(k(t) = k) M\{\Delta V_i(t, \Delta t)/k(t)\} =$$

$$= \Bigg[\lambda p_{0i} a_{0i} + c_i - \left(p_{i0} b_{i0} + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} p_{ij} b_{ij} \right) \sum_k P(k(t) = k) \mu_i(k_i(t)) u(k_i(t)) -$$

$$- \left(q_{i0} \overline{H}_{i0} + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} q_{ij} \overline{H}_{ij} \right) \sum_k P(k(t) = k) \theta_i(k_i(t)) u(k_i(t)) +$$

$$+ \sum_{\substack{j=1 \\ (j \neq i)}}^{n} \sum_k P(k(t) = k) \Big(\mu_j(k_j(t)) u(k_j(t)) p_{ji} a_{ji} +$$

$$+ \theta_j(k_j(t)) u(k_j(t)) q_{ji} \overline{h}_{ji} \Big) \Bigg] \Delta t + o(\Delta t). \qquad (3)$$

Пусть система S_i содержит m_i идентичных линий обслуживания, в каждой из которых время обслуживания заявок распределено по показательному закону с параметром μ_i, $i = \overline{1, n}$. В этом случае

$$\mu_i(k_i(t)) = \begin{cases} \mu_i k_i(t), & k_i(t) \leq m_i, \\ \mu_i m_i, & k_i(t) > m_i, \end{cases} \qquad \mu_i(k_i(t)) u(k_i(t)) = \mu_i \min(k_i(t), m_i).$$

И пусть также

$$\theta_i(k_i(t)) = \begin{cases} \theta_i k_i(t), & k_i(t) \leq m_i, \\ \theta_i m_i, & k_i(t) > m_i, \end{cases} \qquad \text{т.е. } \theta_i(k_i(t)) u(k_i(t)) = \theta_i \min(k_i(t), m_i).$$

Будем предполагать, что усреднение выражения $\mu_i(k_i(t)) u(k_i(t))$ дает величину $\mu_i \min(N_i(t), m_i)$, где $N_i(t) = M\{k_i(t)\}$ - среднее число заявок (ожидающихся и обслуживающихся) в системе S_i, в момент времени t, т.е.

$$M \min(k_i(t), m_i) = \min(N_i(t), m_i), \quad i = \overline{1, n}. \qquad (4)$$

Введем обозначение $v_i(t) = M\{V_i(t)\}$, $i = \overline{1, n}$. Тогда из (1) следует

$$v_i(t + \Delta t) = v_i(t) + M\{\Delta V_i(t, \Delta t)\}.$$

Учитывая (3) и переходя к пределу при $\Delta t \to 0$, имеем неоднородное линейное ОДУ первого порядка

$$\frac{dv_i(t)}{dt} = - \left[\mu_i \left(p_{i0} b_{i0} + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} p_{ij} b_{ij} \right) + \theta_i \left(q_{i0} \overline{H}_{i0} + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} q_{ij} \overline{H}_{ij} \right) \right] \min(N_i(t), m_i) +$$

$$+ \sum_{\substack{j=1 \\ (j \neq i)}}^{n} \left(\mu_j p_{ji} a_{ji} + \theta_j q_{ji} \overline{h}_{ji} \right) \min(N_j(t), m_j) + \lambda p_{0i} a_{0i} + c_i.$$

Задав начальные условия $v_i(0) = v_{i0}$, $i = \overline{1,n}$, можно найти ожидаемые доходы систем сети:

$$v_i(t) = v_{i0} + (c_i + \lambda p_{0i} a_{0i})t +$$

$$+ \sum_{\substack{j=1 \\ (j \neq i)}}^{n} (\mu_j p_{ji} a_{ji} + \theta_j q_{ji} \overline{h}_{ji}) \int_0^t \min(N_j(s), m_j) ds - \left[\mu_i \left(p_{i0} b_{i0} + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} p_{ij} b_{ij} \right) + \right.$$

$$\left. + \theta_i \left(q_{i0} \overline{H}_{i0} + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} q_{ij} \overline{H}_{ij} \right) \right] \int_0^t \min(N_i(s), m_i) ds, \quad i = \overline{1,n}. \tag{5}$$

Замечание. Функция $y = \min(x_i, m_i)$ *является выпуклой кверху и поэтому из неравенства Иенсена следует, что* $M \min(k_i(t), m_i) \leq \min(N_i(t), m_i)$, *причем равенство достигается, когда*

$$N_i(t) = M\{k_i(t)\} = k_i(t), i = \overline{1,n}. \tag{6}$$

Таким образом, при выполнении условия (6) *соотношения* (4), (5) *являются точными. Это условие выполняется, например, когда входящие в сеть потоки заявок являются регулярными, а времена обслуживания заявок в системах постоянными. Отметим также, что условие* (4) *выполняется, когда все СМО сети функционируют в условиях малой нагрузки, т.е.* $\forall t \; k_i(t) \leq m_i$ *или высокой нагрузки, т.е.* $\forall t \; k_i(t) > m_i$, $i = \overline{1,n}$.

ЛИТЕРАТУРА

1. *Статкевич С. Э., Маталыцкий М. А.* Об одном методе исследования НМ-сетей с ограниченным временем ожидания заявок // Вестник Гродненского государственного университета. Сер. 2, 2010. № 1. С. 13–19.

ASYMPTOTIC ANALYSIS OF MARKOV QUEUEING NETWORK WITH UNRELIABLE SYSTEMS

S. Statkevich, T. Rusilko

Grodno State University of Y. Kupala
Grodno, Belarus
sstat@grsu.by

The closed exponentional queueing network with unreliable systems with the large number of messages is investigated. We have received the systems of differential equations for average number of messages and serviceable channels of network systems.

Keywords: unreliable queueing systems, approximation.

1. INTRODUCTION

Let us examine the closed exponential queueing network with the K messages of the same type which consist of $n+1$ queueing systems (QS) $S_0, S_1, \ldots, S_n$. The system S_i includes m_i identical service channels, $i = \overline{1,n}$, and $m_0 = K$.

Considering that service channels of the system S_0 are absolutely reliable and in the other systems systems $S_1, S_2, \ldots, S_n$ the service channels are exposed to random failure; besides the time of the proper functionality of each S_i system's channel has the exponential distribution with the parameter β_i, $i = \overline{1,n}$. After the breakage the channel starts to reconstruct immediately. The time of reconstruction also has the exponential distribution with the parameter γ_i, $i = \overline{1,n}$. After servicing in system S_i the message immediately transfer into the system S_j with probability p_{ij}, $i,j = \overline{0,n}$, $p_{00} = 0$, $\sum\limits_{i=0}^{n} p_{ij} = 1$. The matrix $P = \|p_{ij}\|_{(n+1)\times(n+1)}$ is transition probability matrix of irreducible Markov chains. If the arrived in the system S_j message finds at least one service channel operable and free from the other messages it is immediately serviced and the time of service is a random variable with the parameter μ_i, $i = \overline{1,n}$. Otherwise the message expects the beginning of service without restriction on duration of waiting. Let's assume that if the service channel would fail while completing some message, then after the restoration the interrupted message will be completed. Disciplines of the message processing in the network systems are FIFO.

Our aim is to receive the system of the differential equations for the average number of messages and serviceable channels in the network QS at the large values of K. It should be noted that the presented technics of the results reception has been offered for the first time in the works [1, 2] for the exponentional networks without the specified features (with reliable QS).

2. THE SYSTEM OF EQUATIONS FOR THE STATES PROBABILITIES

Assuming that the service time of messages, durations of serviceable work of channels and restoration time of service channels are independent random variables. The state of such network at the moment t could be described through vector

$$z(t) = (d(t), k(t)) = (d_1(t), d_2(t), \ldots, d_n(t), k_1(t), k_2(t), \ldots, k_n(t)), \tag{1}$$

where $d_i(t)$ and $k_i(t)$ are the numbers of serviceable channels and the messages numbers in the system S_i at the moment t accordingly, $0 \le d_i(t) \le m_i$, $0 \le k_i(t) \le K$, $t \in [0, +\infty)$. It is obvious that $k_0(t) = K - \sum_{i=1}^{n} k_i(t)$ is the number of messages in the system S_0 at the moment t.

Vector $z(t)$ describes $2n-$dimensional Markov process with the continuous time and the definite number of states. Let's consider, that

$$P(d, k, t) = P(d(t) = d, k(t) = k),$$

where $d = (d_1, d_2, \ldots, d_n)$, $0 \le d_i \le m_i$ and $k = (k_1, k_2, \ldots, k_n)$, $0 \le k_i \le K$, $i = \overline{1, n}$. Let's denote I_i as $n-$vector with zero components excluding i, that is equals to 1. Let's describe the possible passages of Markov process $z(t)$ in the state $z(t+\Delta t) = (d, k, t+\Delta t)$ at the time Δt:

- from the state $(d, k + I_i - I_j, t)$ the passage is possible with the probability

$$\mu_i p_{ij} \min(d_i(t), k_i(t) + 1)\Delta t + o(t), \quad i, j = \overline{1, n};$$

- from the state $(d, k - I_i, t)$ with the probability

$$\mu_i p_{ij} \left(K - \sum_{i=1}^{n} k_i(t) + 1 \right) \Delta t + o(t), \quad i = \overline{1, n};$$

- from the state $(d, k + I_i, t)$ with the probability

$$\mu_i p_{i0} \min(d_i(t), k_i(t) + 1)\Delta t + o(t), \quad i = \overline{1, n};$$

- from the state $(d - I_i, k, t)$ with the probability

$$\gamma_i(m_i - d_i(t) + 1)\Delta t + o(t), \quad i = \overline{1, n};$$

- from the state $(d + I_i, k, t)$ with the probability

$$\beta_i(d_i(t) + 1)\Delta t + o(t), \quad i = \overline{1, n};$$

- from the state (d, k, t) with the probability

$$1 - \left[\mu_0 \left(K - \sum_{i=1}^{n} k_i(t) \right) + \sum_{i=1}^{n} \mu_i \min(d_i(t), k_i(t)) + \right.$$
$$\left. + \sum_{i=1}^{n} \gamma_i(m_i - d_i(t)) + \sum_{i=1}^{n} \beta_i d_i(t) \right] \Delta t + 0(\Delta t);$$

• from all other states with the probability $o(\Delta t)$.

Then, the usage of the formula of total probability makes it possible to write the system of difference equations for the probabilities of states from which at $\Delta t \to 0$ we receive the system of difference-differential equations of Kolmogorov for the states probabilities

$$\frac{dP(d,k,t)}{dt} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i p_{ij} \min(d_i(t), k_i(t)) \left[P(d, k - I_i + I_j, t) - P(d,k,t)\right] +$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i p_{ij} \left[\min(d_i(t), k_i(t)+1) - \min(d_i(t), k_i(t))\right] P(d, k - I_i + I_j, t) +$$

$$+ \mu_0 \left(K - \sum_{i=1}^{n} k_i(t)\right) \left[P(d, k - I_j, t) - P(d,k,t)\right] + \mu_0 P(d, k - I_j, t) +$$

$$+ \sum_{i=1}^{n} \mu_i p_{i0} \min(d_i(t), k_i(t)) \left[P(d, k + I_i, t) - P(d,k,t)\right] +$$

$$+ \sum_{i=1}^{n} \mu_i p_{i0} \left[\min(d_i(t), k_i(t)+1) - \min(d_i(t), k_i(t))\right] P(d, k + I_i, t) +$$

$$+ \sum_{j=1}^{n} \gamma_i (m_i - d_i(t)) \left[P(d - I_j, k, t) - P(d,k,t)\right] + \sum_{i=1}^{n} \gamma_i P(d - I_i, k, t) +$$

$$+ \sum_{i=1}^{n} \beta_i d_i(t) \left[P(d + I_i, k, t) - P(d,k,t)\right] + \sum_{i=1}^{n} P(d + I_i, k, t). \qquad (2)$$

The solution of this system in the analytical form is generally inconvenient. Therefore we will consider the important case of the large number of messages in the network, $K >> 1$. In order to determine probability distribution of the random vector $z(t)$, it is convenient to switch to the relative variables, considering vector

$$\xi(t) = \left(\frac{d_1(t)}{K}, \frac{d_2(t)}{K}, \ldots, \frac{d_n(t)}{K}, \frac{k_1(t)}{K}, \frac{k_2(t)}{K}, \ldots, \frac{k_n(t)}{K}\right),$$

In this case possible values of this vector at the fixed t will belong to the bounded closed set

$$G = \left\{(y,k) = (y_1, y_2, \ldots, y_n, x_1, x_2, \ldots, x_n) : x_i \geq 0, \sum_{i=1}^{n} x_i \leq 1, 0 \leq y_i \leq \frac{m_i}{K}\right\} \qquad (3)$$

in which they place in the nodes of the $2n-$dimensional grid at the distance $\varepsilon = \dfrac{1}{K}$ from each other. While magnifying K "the charging density" of the multiple G with the possible components of vector $\xi(t)$ will increase, and it is possible to consider, that it has

a continuous distribution with the probabilities density $p(y,x,t)$, and $K^{2n}P(d,k,t) \to p(y,x,t)$ if $K \to \infty$. Therefore it is possible to use the approximation of the function $P(d,k,t)$, using the relation $K^{2n}P(d,k,t) = K^{2n}P(yK,xK,t) = p(y,x,t)$, $(y,x) \in G$.

Let denote that $e_i = \varepsilon I_i$, $i = \overline{1,n}$, $c(b) = \begin{cases} 1, & b > 0 \\ 0, & b \le 0 \end{cases}$, and

$$\min(b, a+1) = \min(b,a) + c(b-a), \quad c(b-a) = \frac{\partial \min(b,a)}{\partial a}, \tag{4}$$

thus $\min(b,a) = \begin{cases} a, & b \ge a \\ b, & b < a \end{cases}$. Using the relative variables $y_i = \dfrac{d_i}{K}$, $x_i = \dfrac{k_i}{K}$, $l_i = \dfrac{m_i}{K}$ for $i = \overline{1,n}$, expression (4) and that at $K \to +\infty$, $\varepsilon \to 0$, system (2) can be written as follows:

$$\begin{aligned}
\frac{\partial p(y,x,t)}{\partial t} = {} & \sum_{i=1}^{n}\sum_{j=1}^{n} K\mu_i p_{ij} \min(y_i,x_i)\left[p(y,x+e_i-e_j,t) - p(y,x,t)\right] + \\
& + \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i p_{ij} \frac{\partial \min(y_i,x_i)}{\partial x_i} p(y,x+e_i-e_j,t) + \\
& + K\mu_0 \left(1 - \sum_{i=1}^{n} x_i\right)\left[p(y,x-e_j,t) - p(y,x,t)\right] + \mu_0 p(y,x-e_j,t) \\
& + \sum_{i=1}^{n} K\mu_i p_{i0} \min(y_i,x_i)\left[p(y,x+e_i,t) - p(y,x,t)\right] + \\
& + \sum_{i=1}^{n} \mu_i p_{i0} \frac{\partial \min(y_i,x_i)}{\partial x_i} p(y,x+e_i,t) + \\
& + \sum_{i=1}^{n} K\gamma_i (l_i - y_i)\left[p(y-e_i,x,t) - p(y,x,t)\right] + \sum_{j=1}^{n} \gamma_i p(y-e_i,x,t) + \\
& + \sum_{i=1}^{n} K\beta_i y_i \left[p(y+e_i,x,t) - p(y,x,t)\right] + \sum_{j=1}^{n} \beta_i p(y+e_i,x,t).
\end{aligned} \tag{5}$$

3. THE SYSTEM OF DE FOR EXPECTED CHARACTERISTICS

Let's present the right part (5) with the accuracy of term ε^2. If $p(y,x,t)$ is twice continuously differentiated at y and x, than

$$p(y, x \pm e_i, t) = p(y,x,t) \pm \varepsilon \frac{\partial p(y,x,t)}{\partial x_i} + \frac{\varepsilon^2}{2} \frac{\partial^2 p(y,x,t)}{\partial x_i^2} + o(\varepsilon^2),$$

$$p(y, x+e_i-e_j, t) = p(y,x,t) + \varepsilon \left(\frac{\partial p(y,x,t)}{\partial x_i} - \frac{\partial p(y,x,t)}{\partial x_j} \right) +$$

$$+\frac{\varepsilon^2}{2}\left(\frac{\partial^2 p(y,x,t)}{\partial x_i^2}-2\frac{\partial^2 p(y,x,t)}{\partial x_i \partial x_j}+\frac{\partial^2 p(y,x,t)}{\partial x_j^2}\right)+o(\varepsilon^2),$$

$$p(y\pm e_i,x,t)=p(y,x,t)\pm\varepsilon\frac{\partial p(y,x,t)}{\partial y_i}+\frac{\varepsilon^2}{2}\frac{\partial^2 p(y,x,t)}{\partial y_i^2}+o(\varepsilon^2),\quad i=\overline{1,n}. \tag{6}$$

Using them and that $\varepsilon K=1$, it is possible to receive that the density $p(y,x,t)$ satisfies with the accuracy within the term ε^2 to the Kolmogorov-Fokker-Planc equation:

$$\frac{\partial p(y,x,t)}{\partial t}=-\sum_{i=1}^{n}\frac{\partial}{\partial y_i}\left(A_i^{(y)}(y)p(y,x,t)\right)-\sum_{i=1}^{n}\frac{\partial}{\partial_i}\left(A_i^{(x)}(y,x)p(y,x,t)\right)+$$

$$+\frac{\varepsilon}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial y_i y_j}\left(B_{ij}^{(y)}(y)p(y,x,t)\right)+\frac{\varepsilon}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i x_j}\left(B_{ij}^{(x)}(y,x)p(y,x,t)\right), \tag{7}$$

where

$$A_i^{(y)}(y)=\gamma_i\left(l_i-y_i\right), i=\overline{1,n}, \tag{8}$$

$$A_i^{(x)}(y,x)=\sum_{j=1}^{n}\mu_j p_{ji}^{*}\min(y_j,x_j)+\mu_0 p_{0i}\left(1-\sum_{i=1}^{n}x_i\right); \tag{9}$$

$$p_{ji}^{*}=\begin{cases} p_{ji}, & j\neq i,\\ p_{ii}-1, & j=i;\end{cases}\qquad B_{ii}^{(y)}(y)=\gamma_i(l_i-y_i)+\beta_i y_i;\quad B_{ij}^{(y)}(y)=0,\quad i\neq j;$$

$$B_{ii}^{(x)}(y,x)=\sum_{j=1}^{n}\mu_j p_{ji}^{**}\min(y_j,x_j)+\mu_0 p_{0i}\left(1-\sum_{i=1}^{n}x_i\right),$$

$$p_{ji}^{**}=\begin{cases} p_{ji}, & j\neq i,\\ 1-p_{ii}, & j=i;\end{cases}\qquad B_{ij}^{(x)}(x)=-\mu_i p_{ij}\min(y_i,x_i),\quad i\neq j,\quad i,j=\overline{1,n}.$$

As the density $p(y,x,y)$ satisfies the Kolmogorov-Fokker-Planc equation and $A_i^{(y)}(y)$, $A_i^{(x)}(x)$ piecewise linear functions on y, x, according to [3], the mathematical expectations $w_i(t)=M\left\{\frac{d_i(t)}{K}\right\}$, $n_i(t)=M\left\{\frac{k_i(t)}{K}\right\}$, $i=\overline{1,n}$, with the accuracy within the terms of infinitesimal order $O(\varepsilon^2)$ are defined from the systems of the equations

$$\frac{dw_i(t)}{dt}=A_i^{(y)}(w_i(t))=\gamma_i(l_i-w_i(t))-\beta_i w_i(t),\quad i=\overline{1,n}, \tag{10}$$

$$\frac{dn_i(t)}{dt}=A_i^{(x)}(w_i(t),n_i(t))=\sum_{j=1}^{n}\mu_j p_{ji}^{*}\min(w_j(t),n_j(t))-$$

$$+\mu_0 p_{0i}\left(1-\sum_{i=1}^{n}n_i(t)\right),\quad i=\overline{1,n}. \tag{11}$$

The right hand sides of system (11) are continuous piecewise linear functions. By segmentation of phase space and obtaining solutions of system (11) in ranges of right hand sides linearity it is possible to solve whole system.

Let $\Omega(t) = \{1, 2, \ldots, n\}$ be set of vector $n(t)$ component indices. Let's divide $\Omega(t)$ into two disjoint sets $\Omega_0(t)$ and $\Omega_1(t)$:

$$\Omega_0(t) = \{i : w_i(t) < n_i(t) \leq 1\}, \quad \Omega_1(t) = \{j : 0 \leq n_j(t) \leq w_j(t)\}.$$

Each partitioning specifies disjoint regions $G_\tau(t)$ in set

$$G(t) = \left\{ n(t) : n_i(t) \geq 0, \sum_{i=1}^{n} n_i(t) \leq 1 \right\},$$

such that:

$$G_\tau(t) = \left\{ n(t) : w_i(t) < n_i(t) < 1, i \in \Omega_0(t); \right.$$

$$\left. 0 \leq n_j(t) \leq w_j(t), j \in \Omega_1(t); \sum_{c=1}^{n} n_c(t) \leq 1 \right\}, \tau = 1, 2, \ldots, 2^n, \quad \bigcup_{\tau=1}^{2^n} G_\tau(t) = G(t).$$

Then system of equations (11) of explicit form is:

$$\frac{dn_i(t)}{dt} = \sum_{0} \mu_j p^*_{ji} w_j(t) + \sum_{1} \mu_j p^*_{ji} n_j(t) +$$

$$+ \mu_0 p_{0i} \left(1 - \sum_{i=1}^{n} n_i(t) \right), i = \overline{1, n}, \tag{12}$$

for each region $G_\tau(t)$.

The solution of uniform system of equations (10), (12) allows obtaining average relative number of messages and serviceable channels at any queueing system of queueing network.

REFERENCES

1. *Medvedev G. A.* Closed queueing systems and their optimization // Proceedings of the USSR Academy of Sciences: Engineering Cybernetics, 1978. № 6. P. 199–203.
2. *Medvedev G. A.* About optimisation of closed queueing systems // Proceedings of the USSR Academy of Sciences: Engineering Cybernetics, 1975. № 6. P. 65–73.
3. *Parajev Y. I.* Introduction statistical dynamics of control processes. M.: Sov. Radio, 1976.

НАХОЖДЕНИЕ ЗАКОНА УПРАВЛЕНИЯ ЦЕНОЙ ПРОДАЖИ ТОВАРА

Н. Степанова

Томский государственный университет
Томск, Россия
natalia0410@rambler.ru

Аннотация: Оптимизация продажи скоропортящейся продукции (молоко, творог и т.д.) представляет определенный практически интерес, так как продукцию, не реализованную в течение торговой сессии, в лучшем случае надо пускать в переработку, а в худшем случае - просто выбрасывать. Поэтому при реализации такой продукции возникает ряд вопросов, таких как а) какой объем продукции надо завозить на торговую точку; б) по какой цене ее продавать; в) как управлять ценой продажи продукции, чтобы к концу торговой сессии она была полностью реализована. Все эти задачи надо решать при вполне естественном критерии оптимальности - максимизации прибыли, получаемой от реализации продукции.

Ключевые слова: интервал, распределение, диффузионное приближение, аппроксимация.

1. ВВЕДЕНИЕ

Пусть в торговую точку завозится партия продукции объемом Q_0, которая должна быть продана в течение торговой сессии длительности . Пусть d - объем затрат на выпуск единицы продукции, так что производителю эта партия стоила $Q_0 d$ рублей.

Пусть $c(t)$ есть цена, по которой продукция продается в момент времени t. В данной работе рассматривается вопрос управления ценой продажи $c(t)$ в зависимости от времени t и объема $Q(t)$ продукции, не реализованной к этому моменту времени. Цель этого управления - добиться того, что продукция будет реализована к концу торговой сессии и при этом будет получена максимальная прибыль.

Будем считать, что поток покупателей является пуассоновским потоком интенсивности $\lambda(c)$, зависящей от розничной цены . Вид этой зависимости будет уточнен ниже.

Будем считать, что покупатели покупают товар независимо друг от друга, и объем покупки ξ есть случайная величина с $M\{\xi\} = \alpha_1$ и $M\{\xi^2\} = \alpha_2$.

Пусть $Q(t)$ есть количество продукции, которая осталась не реализованной в момент времени t. Рассмотрим решение задачи в так называемом диффузионном приближении, когда $Q(t)$ аппроксимируется диффузионным случайным процессом.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть теперь производится управление ценой продажи товара, и цена продажи $c(t)$ в момент времени t выбирается из условия

$$a_1\lambda(c) = \frac{Q(t)}{\varphi(t)}, \tag{1}$$

с некоторой, пока неизвестной функцией $\varphi(t)$.

Тогда, в детерминированном приближении, $Q(t)$ определяется решением следующего дифференциального уравнения

$$\frac{\partial Q}{\partial t} = -\mu Q(t) - \frac{Q(t)}{\varphi(t)}, \tag{2}$$

полученного подстановкой, которое надо решить при начальном условии

$$Q(0) = Q_0.$$

Разделяя переменные и интегрируя, получим

$$Q(t) = Q_0 exp\left(-\mu t - \int_0^t \frac{\partial z}{\varphi(z)}\right) \tag{3}$$

Рассмотрим частный случай, когда зависимость $\lambda(c)$ имеет вид

$$\lambda(c) = \lambda_0 - \lambda_1 \frac{c - c_0}{c_0}.$$

Тогда цена продажи $c(t)$ в момент времени t находится из условия

$$\alpha_1\lambda(c) = \alpha_1\left(\lambda_0 + \lambda_1 - \lambda_1\frac{c}{c_0}\right) = \frac{Q(t)}{\varphi(t)},$$

откуда получаем

$$c(t) = c_0\left(1 + \frac{\lambda_1}{\lambda_0} - \frac{Q(t)}{\lambda_1\alpha_1\varphi(t)}.\right) \tag{4}$$

Выручка от продажи нашей партии товара в течение времени T будет равна

$$S = \int_0^T c(t)\alpha_1\lambda(c(t))dt = c_0\left(1 + \frac{\lambda_1}{\lambda_0}\right)\int_0^T \frac{Q(t)}{\varphi(t)}dt - \frac{c_0}{\alpha_1\lambda_1}\int_0^T \frac{Q^2(t)}{\varphi^2(t)}dt. \tag{5}$$

Если товар приобретается по оптовой цене d, то прибыль от его продажи будет равна

$$P = S - dQ_0 = c_0\left(1 + \frac{\lambda_1}{\lambda_0}\right)\int_0^T \frac{Q(t)}{\varphi(t)}dt - \frac{c_0}{\alpha_1\lambda_1}\int_0^T \frac{Q^2(t)}{\varphi^2(t)}dt. - dQ_0 \tag{6}$$

Найдем оптимальный вид функции $\varphi(t)$. Мы видим, что зависит от двух функционалов

$$\int_0^T \frac{Q(t)}{\varphi(t)}dt \quad \text{и} \quad \int_0^T \frac{Q^2(t)}{\varphi^2(t)}dt.$$

Обозначая $Q(t)/\varphi(t) = f(t)$ получим зависимость от двух функционалов -

$$\int_0^T f(t)dt \quad \text{и} \quad \int_0^T f^2(t)dt.$$

В любом случае, попытка найти $max\, P$ по виду функции $f(t)$ приводит к задаче вида

$$\int_0^T f^2(t)dt + \kappa \int_0^T f(t)dt \Rightarrow extr, \tag{7}$$

где κ - неопределённый множитель Лагранжа. Приравнивая нулю вариацию по $f(t)$, получим, что $f(t) = const$.

Таким образом,

$$f(t) = \frac{Q(t)}{\varphi(t)} = C,$$

или, в явном виде,

$$\frac{1}{\varphi(t)} exp\left(-\mu t - \int_0^t \frac{dz}{\varphi(z)}\right) = C. \tag{8}$$

Переписывая его в виде

$$\frac{1}{\varphi(t)} exp\left(-\int_0^t \frac{dz}{\varphi(z)}\right) = Ce^{\mu^t}$$

и замечая, что

$$\frac{1}{\varphi(t)} exp\left(-\int_0^t \frac{dz}{\varphi(z)}\right) = -\left(exp\left(-\int_0^t \frac{dz}{\varphi(z)}\right)\right)^t,$$

получим

$$\left(exp\left(-\int_0^t \frac{dz}{\varphi(z)}\right)\right)^t = -Ce^{\mu^t}.$$

Отсюда имеем

$$exp\left(-\int_0^t \frac{dz}{\varphi(z)}\right) = -\frac{C}{\mu}e^{\mu^t} + C_1.$$

Так как при $t = 0 \quad \int_0^t \frac{dz}{\varphi(z)} = 0$, то

$$1 = -\frac{C}{\mu} + C_1, \quad C_1 = 1 + \frac{C}{\mu},$$

и поэтому

$$exp\left(-\int_0^t \frac{dz}{\varphi(z)}\right) = \frac{C}{\mu} - \frac{C}{\mu}e^{\mu^t} + 1.$$

Логарифмируя это выражение и дифференцируя, получим, после некоторых упрощений

$$\varphi(t) = \frac{\widetilde{C}e^{-\mu^t} - 1}{\mu}.$$

Возьмем константу $\widetilde{C}$ в виде $\widetilde{C} = e^{\mu CT}$ с некоторым $C \geq 1$. Тогда окончательно закон управления ценой примет вид

$$\alpha_1 \lambda(c) = \frac{Q(t)}{\varphi(t)}, \quad \varphi = \frac{e^{\mu(CT-t)} - 1}{\mu}. \tag{9}$$

Именно этот закон управления ценой и будет рассматриваться в дальнейшем. Заметим, что при $\mu \to 0 \quad \varphi(t) \to CT - t$, то есть мы получаем тот вид $\varphi(t)$, который обеспечивал максимум прибыли при продаже скоропортящихся товаров в предыдущих разделах.

Найдём теперь выручку и прибыль при данном законе управления ценой в детерминированном случае. Вычисляя интегралы, получим

$$Q(t) = Q_0 \frac{e^{\mu(CT-t)} - 1}{e^{\mu CT} - 1},$$

$$\int_0^T \frac{Q(t)}{\varphi(t)} dt = \frac{Q_0 \mu T}{e^{\mu CT} - 1}, \quad \int_0^T \frac{Q^2(t)}{\varphi^2(t)} dt = \frac{Q_0^2 \mu^2 T}{(e^{\mu CT} - 1)^2}.$$

Поэтому прибыль от продажи нашей партии товара равна

$$P = c_0 \left(1 + \frac{\lambda_0}{\lambda_1}\right) \frac{Q_0 \mu T}{e^{\mu CT} - 1} - \frac{c_0}{\lambda_1 \alpha_1} \frac{Q_0^2 \mu^2 T}{(e^{\mu CT} - 1)^2} - dQ_0. \tag{10}$$

Отсюда находится оптимальное значение Q_0:

$$Q_0 = \left[c_0 \left(1 + \frac{\lambda_0}{\lambda_1}\right) \frac{\mu T}{e^{\mu CT} - 1} - d\right] \frac{\lambda_1 \alpha_1 (e^{\mu CT} - 1)^2}{2\mu^2 T c_0}, \tag{11}$$

и максимальное значение прибыли

$$P_{max} = \left[c_0 \left(1 + \frac{\lambda_0}{\lambda_1}\right) \frac{\mu T}{e^{\mu CT} - 1} - d\right]^2 \frac{\lambda_1 \alpha_1 (e^{\mu CT} - 1)^2}{4\mu^2 T c_0}. \tag{12}$$

3. ЗАКЛЮЧЕНИЕ

Однако в этом случае есть дополнительная возможность - провести оптимизацию и по параметру . Обозначая комбинацию $e^{\mu Ct} - 1 = z$, получим

$$P_{max} = \left[c_0 \left(1 + \frac{\lambda_0}{\lambda_1}\right) \mu T z - dz^2\right]^2 \frac{\lambda_1 \alpha_1}{4\mu^2 T c_0},$$

и оптимальное значение z равно

$$z_{opt} = \frac{c_0 (1 + \lambda_0/\lambda_1) \mu T}{2d} = e^{\mu C_{opt} T} - 1.$$

Отсюда оптимальное значение есть

$$C_{opt} = \frac{1}{\mu T} ln \left(1 + \frac{c_0(1 + \lambda_0/\lambda_1)\mu T}{d}\right). \tag{13}$$

Так как $C \geq 1$, то окончательно

$$C_{opt} = max \left(1, \frac{1}{\mu T} ln \left(1 + \frac{c_0(1 + \lambda_0/\lambda_1)\mu T}{d}\right)\right).$$

ЛИТЕРАТУРА

1. *Степанова Н.В., Терпугов А.Ф.* Оптимальное управление ценой при продаже скоропотящегося товара // Вестник Сибирского государственного аэрокосмического университета, 2007, вып. 4(17), с. 35-39.
2. *Степанова Н.В., Терпугов А.Ф.* Управление ценой при продаже скоропортящейся продукции // Вестник Томского госуниверситета. Сер. Управление, вычислительная техника и информатика. 2007. №1. с. 22-35.
3. *Степанова Н.В.* Управление ценой при продаже портящегося товара // Научное творчество молодежи. Материалы XII Всероссийской научно-практической конференции. Анжеро-Судженск, 18-19 апреля 2008 г. Издательство Томского университета, 2008. Ч. 1, с. 40-43.

PROCESSOR SHARING SYSTEMS WITH RANDOM CAPACITY DEMANDS

O. Tikhonenko

Czestochowa University of Technology, Institute of Mathematics
Czestochowa, Poland
oleg.tikhonenko@im.pcz.pl

We discuss processor sharing queueing systems with non-homogeneous demands. This non-homogenity means that each demand (independently of others) has some random capacity and its length (or amount of work for its service) generally depends on the capacity. In real systems, a total sum of capacities of demands presenting in the system is limited by some constant value (memory volume) $V > 0$. But we estimate loss characteristcs for such system using queueing models with unlimited memory volume.

Keywords: random capacity demand, total demands capacity, memory volume.

1. INTRODUCTION

Egalitarian processor sharing (EPS) systems are used for modeling of computer and communicating networks [1]. Presently, they are applicable to situations where a common resource is shared by a varying number of concurrent users [2] (for example, to WEB-servers modeling [3]).

We introduce the following additional assumption for the classical $M/G/1-EPS$ system. Assume that each demand is characterized by some non-negative random capacity. This random variable can be interpreted as a part of system's memory space used by the demand during its presence in the system. A total sum of demands capacities $\sigma(t)$ in the system at arbitrary time instant t is referred as the total demands capacity. The random value $\sigma(t)$ can be limited by some constant value V $(0 < V < \infty)$, which is called the memory volume of the system. In this case we have a non-classical processor sharing system that will be notated by $M/G/1(V)-EPS$. Later on, we shall call demand length the amount of work necessary for demand's service, i.e. the service time under condition that there are no other demands in the system during its presence in it. Analogously, we shall call residual length of the demand its residual service time after some time instant under the same condition (see [2]).

The purpose of the paper is 1) to obtain the non-stationary and stationary distribution of total demands capacity in the system $M/G/1-EPS$; 2) to determine some estimations of loss characteristics for systems $M/G/1(V)-EPS$ with limited memory volume $(V < \infty)$ based on the model with unlimited one; 3) to compare processor sharing systems $M/G/1(V)-EPS$ and $M/G/1-EPS$ from the point of view of estimation of loss characteristics.

2. CLASSICAL PROCESSOR SHARING SYSTEM

Denote by $\eta(t)$ the number of demands present in the system at the time instant t and $\xi_i^*(t)$ be the residual length of ith demand at this instant, $i = \overline{1, \eta(t)}$. Let $F(x,t) = \mathbf{P}\{\zeta < x, \xi < t\}$ be the joint distribution function of the demand capacity ζ and its length ξ (we assume that the demand capacity and its length doesn't depend on his arrival time and on characteristics of other demands). Then $L(x) = F(x,\infty)$ and $B(t) = F(\infty,t)$ be the distribution functions of the random variables ζ and ξ consequently. Let a be an arrival rate of entrance flow of demands, $\alpha(s,q) = \int_0^\infty \int_0^\infty e^{-sx-qt} dF(x,t)$ be the double Laplace-Stieltjes transform (with respect to x and t) of the distribution function $F(x,t)$, $\varphi(s) = \alpha(s,0)$ and $\beta(q) = \alpha(0,q)$ be the Laplace-Stieltjes transform (LST) of the distribution functions $L(x)$ and $B(t)$ consequently, $D(x,t) = \mathbf{P}\{\sigma(t) < x\}$ be the distribution function of total demands capacity at the time instant t, $\delta(s,t) = \int_0^\infty e^{-sx} d_x D(x,t)$ be the LST of the function $D(x,t)$ with respect to x, $\overline{\delta}(s,q) = \int_0^\infty e^{-qt}\delta(s,t)dt$ be the Laplace transform of the function $\delta(s,t)$ with respect to t. The mixed $(i+j)$th moments of the random variables ζ and ξ (if they exist) take the form: $\alpha_{i\,j} = (-1)^{i+j} \frac{\partial^{i+j}}{\partial s^i \partial q^j} \alpha(s,q)\Big|_{s=0,q=0}$.

Assume that demands in the considered system at an arbitrary time t are numerated as random; i.e. if the number of demands is k, then there are $k!$ ways to enumerate them, and each enumeration can be chosen with the same probability $1/k!$.

One can easily show that the system under consideration is described by the Markov process

$$(\eta(t), \xi_i^*(t), i = \overline{1, \eta(t)}), \tag{1}$$

where components $\xi_i^*(t)$ are absent if $\eta(t) = 0$. In this case we also have $\sigma(t) = 0$.

In what follows, to simplify the notation, we denote $Y_k = (y_1, \ldots, y_k)$. We characterize the process (1) by functions with the following probabilistic sense:

$$P_0(t) = \mathbf{P}\{\eta(t) = 0\}; \tag{2}$$

$$\Theta_k(Y_k, t) = \mathbf{P}\{\eta(t) = k, \xi_j^*(t) < y_j, j = \overline{1,k}\},\ k = 1, 2, \ldots; \tag{3}$$

$$P_k(t) = \mathbf{P}\{\eta(t) = k\} = \Theta_k(\infty_k, t),\ k = 1, 2, \ldots, \tag{4}$$

where $\infty_k = (\infty, \ldots, \infty)$ is a k-component vector.

Note that the functions $\Theta_k(Y_k, t)$ are symmetric with respect to permutations of components of the vector Y_k due to our random enumeration of customers in the system.

Let us determine the function $\overline{\delta}(s,q)$ under zero initial condition $\eta(0) = \sigma(0) = 0$.

Denote by $\overline{p}_0(q) = \int_0^\infty e^{-qt} P_0(t)dt$ and $\overline{\theta}_k(Y_k, q) = \int_0^\infty e^{-qt}\Theta_k(Y_k, t)dt$ the Laplace transforms with respect to t of the functions $P_0(t)$ and $\Theta_k(Y_k, t)$ consequently. It's known (see [2]) that

$$\overline{p}_0(q) = [q + a - a\pi(q)]^{-1} \tag{5}$$

under zero initial condition, where $\pi(q)$ is the LST of the busy period distribution function for the system under consideration. Note [2] that $\pi(q)$ is a unique solution of the functional equation $\pi(q) = \beta(q + a - a\pi(q))$ such that $|\pi(q)| \le 1$.

Lemma. *Under zero initial condition, the functions* $\overline{\theta}_k(Y_k, q)$, $k = 1, 2, \ldots$, *have the form* $\overline{\theta}_k(Y_k, q) = \overline{p}_0(q) \prod_{i=1}^{k} \int_0^{y_i} [q + a - aB(u)]du$.

Let $\beta_i = \mathbf{E}\xi^i = (-1)^i \beta^{(i)}(0)$ be the ith moment of customer length, $i = 1, 2, \ldots$.

Corollary 1. *If* $\rho = a\beta_1 < 1$, *limits* $\theta_k(Y_k) = \lim_{t\to\infty} \Theta_k(Y_k, t)$, $k = 1, 2, \ldots$, *exist being independent of initial condition and have the form* $\theta_k(Y_k) = (1-\rho)a^k \prod_{i=1}^{k} \int_0^{y_i} [1 - -B(u)]du$.

Corollary 2. *Let* $\overline{p}_k(q)$ *be the Laplace transform of the function* $P_k(t)$, $k = 0, 1, \ldots$, *under zero initial condition. Then we have* $\overline{p}_k(q) = \frac{a^k(1-\pi(q))^k}{(q+a-a\pi(q))^{k+1}}$.

From the corollary 1 we can obtain the known relation for the stationary distribution $\{p_k\}$ of the number of demands in the system ($\rho = a\beta_1 < 1$)[2]: $p_k = \theta_k(\infty_k) = = (1-\rho)\rho^k$, $k = 0, 1, \ldots$.

Let $\chi(t)$ be the capacity of a demand being on service at the time t and $\xi^*(t)$ be the residual length of this demand at the time t. We shall use the notation $E_y(x) = = \mathbf{P}\{\chi(t) < x | \xi^*(t) = y\}$. It is known [4] that the LST of the conditional distribution function $E_y(x)$ has the form:

$$e_y(s) = [1 - B(y)]^{-1} \int_{x=0}^{\infty} e^{-sx} \int_{u=y}^{\infty} dF(x, u). \tag{6}$$

Theorem 1. *For zero initial condition the function* $\overline{\delta}(s, q)$ *is determined by the relation* $\overline{\delta}(s, q) = \{[q+a-a\pi(q)][1-I(s, q)]\}^{-1}$, *where* $I(s, q) = \int_0^{\infty} (q+a-aB(y))e_y(s)dy$ *and* $e_y(s)$ *is determined by the relation* (6).

Corollary 3. *If the random variables* ζ *and* ξ *are independent, we have:*

$$\overline{\delta}(s, q) = [q + a(1 - \pi(q))(1 - \varphi(s))]^{-1}. \tag{7}$$

Corollary 4. *Under zero initial condition, the Laplace transform* $g(s, q)$ *with respect to* t *of generation function* $P(z, t) = \sum_{k=0}^{\infty} P_k(t)z^k$, $|z| \le 1$, *of the demands number in the system at time instant* t *have the following form:*

$$g(z, q) = \int_0^{\infty} e^{-qt} P(z, t)dt = [q + a(1 - z)(1 - \pi(q))]^{-1}. \tag{8}$$

Corollary 5. *Let* $\rho = a\beta_1 < 1$. *Then stationary mode exists. The LST* $\delta(s)$ *of the stationary distribution function* $D(x) = \lim_{t\to\infty} D(x, t)$ *of demands total capacity has the form:*

$$\delta(s) = \frac{1 - \rho}{1 + a\alpha'_q(s, q)|_{q=0}}. \tag{9}$$

Note that the relation (9) was first obtained by B. Sengupta [5].

Corollary 6. *Let* $\delta_1(t)$ *be the first moment of the total demands capacity* $\sigma(t)$ *under zero initial condition,* $\overline{\delta}_1(q)$ *be the Laplace transform of the function* $\delta_1(t)$. *Then we have:*

$$\overline{\delta}_1(q) = \frac{a\alpha_{11} + q \int_0^{\infty} \int_0^{\infty} xS(t)dF(x, t)}{[q + a - a\pi(q)] \left[1 - \rho - q \int_0^{\infty} S(t)dB(t)\right]^2},$$

where $S(t) = \int_0^t [1 - B(y)]^{-1} dy$.

Let σ be a stationary total demands capacity ($\sigma(t) \Rightarrow \sigma$ in the sense of a weak convergence). From the relation (9) the following known formulas [5] can be obtained:

$$\delta_1 = \mathbf{E}\sigma = -\delta'(0) = \frac{a\alpha_{11}}{1-\rho}, \ \delta_2 = \mathbf{E}\sigma^2 = \delta''(0) = \frac{a\alpha_{21}}{1-\rho} + 2\delta_1^2. \tag{10}$$

For some special cases we can obtain the view of the distribution function $D(x)$ from the formula (9). For example, consider the case when the demand capacity ζ and its length ξ are connected by the relation $\xi = c\zeta + \xi_1$, $c > 0$, where the random variables ζ and ξ_1 are independent (such dependence for demand capacity and its length is true for many real information systems).

Denote by $\kappa_1 = \mathbf{E}\xi_1$ the first moment of the random variable ξ_1. In this case we have $\alpha(s,q) = \varphi(s+cq)\kappa(s)$, where $\kappa(s)$ is the Laplace–Stieltjes transform of the distribution function of the random variable ξ_1. The the relation (9) takes the following form:

$$\delta(s) = \frac{1-\rho}{1 + a[c\varphi'(s) - \kappa_1\varphi(s)]}. \tag{11}$$

Assume that customer capacity ζ has an exponential distribution with the parameter $f > 0$. Then from the formula (11) we obtain: $\delta(s) = \frac{(1-\rho)(s+f)^2}{(s+f)^2 - \rho_1 f^2 - \rho_2 f(s+f)}$, where $\rho_1 = ac/f$, $\rho_2 = a\kappa_1$, so that $\rho = a\beta_1 = \rho_1 + \rho_2$.

Now we can determine the original of Laplace transform $\delta(s)/s$, where $\delta(s)$ is defined by formula (11), and obtain the view of the stationary distribution function $D(x)$:

$$D(x) = 1 - \frac{(1-\rho)e^{-fx}}{2b}\left[\frac{(\rho_2+b)^2 e^{(\rho_2+b)fx/2}}{2-\rho_2-b} - \frac{(\rho_2-b)^2 e^{(\rho_2-b)fx/2}}{2-\rho_2+b}\right], \tag{12}$$

where $b = \sqrt{\rho_2^2 + 4\rho_1}$.

3. ESTIMATION OF LOSS CHARACTERISTICS

The $M/G/1-EPS$ is a system without losing of customers ($V = \infty$). But with the help of this model we can estimate the memory capacity V in order to guarantee inexceeding of given loss probability.

Assume that we have a stationary queueing system Q_∞ with Poisson entrance flow without losses of demands. Let Q_V be a stationary system that differs from Q_∞ only with the fact that its total capacity is limited by the constant value V. We denote by $D(x)$ the distribution function of total demands capacity for the system Q_∞ and by $D_V(x)$ the distribution function of this random value for the system Q_V.

Theorem 2. *The inequality* $D(x) \le D_V(x)$ *takes place for all* $x > 0$.

Proof of the theorem see in [4].

It follows from theorem 2 that the loss probability P for the system Q_V satisfies the following inequality [4]:

$$P = 1 - \int_0^V D_V(V-x)dL(x) \le 1 - \int_0^V D(V-x)dL(x) = P^*. \tag{13}$$

Thus, the value P^* is an upper estimation of loss probability for the system Q_V. If we choose V under condition that P^* is given so that the equality $\int_0^V D(V-x)dL(x) = 1 - P^*$ is satisfied, then the real loss probability P doesn't exceed P^*. If only very rare losses are permitted in the system under consideration, the difference between the values P and P^* is inessential.

Note that the loss probability is not exhaustive characteristic of losses, because its value shows a part of losing demands, not a part of losing capacity or, in other words, information being lost. Really, it is obvious that demands having large capacity will be lost more often. Therefore, more objective losses estimation is the value $Q = 1 - \frac{1}{\varphi_1}\int_0^V xD_V(V-x)dL(x)$.

The value Q is the probability of losing of a unit of demand capacity. The next inequality follows from theorem 2:

$$Q = 1 - \frac{1}{\varphi_1}\int_0^V xD_V(V-x)dL(x) \le 1 - \frac{1}{\varphi_1}\int_0^V xD(V-x)dL(x) = Q^*.$$

If only very rare losses are permitted in the system under consideration, the difference between the values Q and Q^* is inessential.

For example, in the case of the distribution function (12) we obtain:

$$P^* = \left\{1 - \frac{1-\rho}{b}\left[a_1\frac{1-e^{-(1-b_1)fV}}{b+\rho_2} + a_2\frac{1-e^{-(1-b_2)fV}}{b-\rho_2}\right]\right\}e^{-fV},$$

where $a_1 = \frac{(\rho_2+b)^2}{2-\rho_2-b}$, $a_2 = \frac{(\rho_2-b)^2}{2-\rho_2+b}$, $b_1 = -1 + \frac{\rho_2+b}{2}$, $b_2 = -1 + \frac{\rho_2-b}{2}$;

$$Q^* = \left\{1 + fV - \frac{2(1-\rho)}{b}\left[\frac{(a_1+a_2)fV}{8\rho_1} + a_1\frac{1-e^{-(1-b_1)fV}}{(b+\rho_2)^2} - a_2\frac{1-e^{-(1-b_2)fV}}{(b-\rho_2)^2}\right]\right\}e^{-fV}.$$

Note that in most cases the calculation and estimation of the probability Q is very complicated. Therefore, we often must restrict ourselve to the calculation and estimationof the loss probability P.

If it is impossible to determine the view of the distribution function $D(x)$, we can estimate the value P^* by approximation of the function $\Phi(x) = \int_0^x D(x-u)dL(u)$, being the distribution function of the sum of independent random variables σ and ζ, with the distribution function of gamma distribution $\Phi^*(x) = \gamma(h, rx)/\Gamma(h)$, where $\gamma(h, rx) = \int_0^{rx} t^{h-1}e^{-t}dt$ is the incomplete gamma function, $\Gamma(h) = \gamma(h, \infty)$ is the gamma function. The parameters h and r of the approximate distribution should be chosen so that its first and second moments $f_1^* = h/r$ and $f_2^* = h(h+1)/r^2$ be equal to the first and second moments of the distribution function $\Phi(x)$ respectively. It is obvious that these moments have the form

$$f_1 = \delta_1 + \varphi_1, \;\; f_2 = \delta_2 + \varphi_2 + 2\delta_1\varphi_1. \tag{14}$$

Thus, the parameters of the distribution function $\Phi^*(x)$ should be chosen as follows: $h = \frac{f_1^2}{f_2 - f_1^2}$, $r = \frac{f_1}{f_2 - f_1^2}$, where f_1 and f_2 can be calculated from relations (10), (14). Hence, we have the approximate formula $P^* \cong 1 - \Phi^*(V)$.

Note that in the case of not very small permissible loss probabilities, using the estimation P^* instead of P leads to unjustifiably surplus choice of the capacity volume V. Therefore, the direct analysis of processor sharing systems with limited memory space is very important.

4. THE CASE OF LIMITED TOTAL CAPACITY

The system $M/G/1(V) - EPS$ with demands of different types was analyzed in detail in [6]. We shall concider a special case of demands of the same type. Then, for stationary probabilities of number of demands present in the system we have:

$$p_0 = \left(\sum_{k=0}^{\infty} a^k A_*^{(k)}(V) \right)^{-1}, \; p_k = p_0 a^k A_*^{(k)}(V), \; k = 1, 2, \ldots,$$

where $A_*^{(k)}(x)$ is a kth order Stieltjes convolution of the function $A(x) = \int_{u=0}^{x} \int_{t=0}^{\infty} u dF(u,t)$. The loss probability has the form

$$P = 1 - p_0 \left[L(V) - \sum_{k=1}^{\infty} a^k A_*^{(k)}(V) \right].$$

Assume additionally that demand capacity has an exponential distribution with parameter f, and let the demand length be proportional to its capacity ($\xi = c\zeta$, $c > 0$). Then, we obtain after some calculation:

$$p_0 = \begin{cases} \dfrac{1-\rho}{1 - \sqrt{\rho} e^{-fV} \left[\sinh(\sqrt{\rho} fV) + \sqrt{\rho} \cosh(\sqrt{\rho} fV) \right]}, & \text{if } \rho \neq 1, \\ \dfrac{1 + e^{-2fV}}{1 + fV}, & \text{if } \rho = 1; \end{cases}$$

$$p_k = p_0 \rho^k \left[1 - e^{-fV} \sum_{i=0}^{2k-1} \frac{(fV)^i}{i!} \right], \; k = 1, 2, \ldots; \; P = p_0 e^{-fV} \cosh(\sqrt{\rho} fV),$$

where $\rho = ac/f$.

Now we can compare values P^* and P or Q^* and Q using analytical results or simulation. Table 1 presents the dependence of loss characteristics upon the memory capacity V. We assume here that $\rho = 0.6$, the demand length is proportional to its capacity ($\xi = c\zeta$), where $c = 1$, and capacity ζ has an exponential distribution with parameter $f = 1$.

Values P^*, Q^*, P were obtained by calculation from above relations, and the value Q was estimated by simulation. The table shows that estimators P^*, Q^* are not very precise and we can use them for the case when the proper loss characteristics are near zero.

Table 1: Probabilities P and Q for $\rho = 0.6$

V	P^*	Q^*	P	Q
0.0	1.00000	1.00000	1.00000	1.00000
0.2	0.92721	0.99569	0.81994	0.98269
0.4	0.86622	0.98366	0.67754	0.94034
0.6	0.81392	0.96529	0.56700	0.88482
0.8	0.76815	0.94194	0.48156	0.82409
1.0	0.72735	0.91487	0.41516	0.76311
2.0	0.56855	0.75562	0.23586	0.51290
3.0	0.45178	0.60242	0.15775	0.35596
4.0	0.35651	0.47628	0.11281	0.25640
5.0	0.28750	0.37679	0.08340	0.18993
6.0	0.22947	0.29888	0.06291	0.14330
7.0	0.18316	0.23763	0.04811	0.10963
8.0	0.14620	0.18925	0.03716	0.08464
10.0	0.09314	0.12034	0.02263	0.05165
15.0	0.03018	0.03896	0.00697	0.01589
20.0	0.00978	0.01262	0.00222	0.00512
30.0	0.00103	0.00133	0.00023	0.00054
40.0	0.00011	0.00014	0.00002	0.01589
50.0	0.00001	0.00002	0.00000	0.00001

REFERENCES

1. *Litjens R., Van der Berg H., Boucherie R. J.* Throughputs in processor sharing models for integrated stream and elastic traffic // Performance Evaluation. 2008. V. 65. P. 152–180.
2. *Yashkov S. F.* Analysis of Queues in Computers. Radio i Svyaz, Moskow, 1989. (In Russian).
3. *Yashkov S. F., Yashkova A. S.* Processor sharing: a survey of the mathematical theory // Autom. Remote Control. 2007. V. 68. № 9. P. 1662–1731.
4. *Tikhonenko O.* Computer systems probability analysis. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2006. (In Polish).
5. *Sengupta B.* The spatial requirement of an $M/G/1$ queue or: how to design for buffer space // Lect. Notes Contr. Inf. Sci. 1984. V. 60. P. 547–564.
6. *Tikhonenko O. M.* Queueing systems with processor sharing and limited resources // Autom. Remote Control. 2010. V. 71. № 9. P. 804–815.

STATISTICAL ANALYSIS OF QUEUEING NETWORKS WITH INFINITE NUMBER OF SERVERS

G. Tsitsiashvili

Institute for Applied Mathematics Far Eastern Branch of RAS
Vladivostok
guram@iam.dvo.ru

In this paper a problem of successive changing of individuals states or technical means from some population is considered. This changing is connected as with individuals aging so with appearance of new individuals. The problem is solved using queueing networks with infinite numbers of servers in their nodes.

Keywords: product theorem, Markov process, stationary distribution.

1. INTRODUCTION

Last years in the reliability theory a large attention is devoted to the life time models with a large number of states [1]. New applications appear in models of a dynamics of populations with multistage individuals [2]. These applications demand to develop theory of queueing networks with infinite number of servers beginning from product theorems through an aggregation of nodes to convenient statistical procedures of parameters estimates.

2. OPENED NETWORKS

Consider opened queueing network with the nodes set $S = \{0\} \bigcup I,\ I = \{1, \ldots, m\}$, the indivisible route matrix $\Theta = ||\theta(i,j)||_{i,j \in S}$ and Poisson input flow with parameter $\lambda(0) > 0$. The node $i \in I$ contains infinite number of servers with exponentially distributed service times with the parameter $\mu(i) > 0$. The node 0 is imaginary, it is a source of arrival customers and a runoff for customers leaving the network. The route matrix Θ is called indivisible if for any $i, j \in S,\ i \neq j$, there are $i_1, \ldots, i_s \in S$, so that the inequality

$$0 < \theta(i, i_1) \cdot \theta(i_1, i_2) \cdot \ldots \cdot \theta(i_{s-1}, i_s) \cdot \theta(i_s, j). \tag{1}$$

is true. From (1) we have that for fixed $\lambda(0) > 0$ the system of linear algebraic equations

$$(\lambda(0), \lambda(1), \ldots, \lambda(m)) = \Theta(\lambda(0), \lambda(1), \ldots, \lambda(m)) \tag{2}$$

[3] has single solution $(\lambda(1), \ldots, \lambda(m)),\ \lambda(i) > 0,\ i \in I$. Then the process $(n_1(t), \ldots, n_m(t))$ which describes numbers of customers in nodes of opened network [4, Proposition

1.10, Example 1.29] is ergodic and its limit distribution

$$P(n_1,\dots,n_m)=\prod_{i=1}^{m}p_i(n_i),\ p_i(n_i)=\mathbf{e}^{(}-\rho(i))\frac{(\rho(i))^{n_i}}{n_i!},\ \rho(i)=\frac{\lambda(i)}{\mu(i)},\ i\in I. \tag{3}$$

Theorem 1. *Almost surely*

$$\lim_{T\to\infty}\frac{\int_0^T n_i(t)dt}{T}=\rho_i,\ i\in I. \tag{4}$$

Proof. From the formula (3) the distribution $p_i(n_i)$ is Poisson and its mean coincides with $\rho(i)$. So the law of large numbers for ergodic markov processes [4, Theorem 1.2] leads to the formula (4). □

Theorem 2. *Suppose that the sets $I_1,\dots,I_r$ create a decomposition of the set I onto nonintersected subsets. Denote $N_k(t)=\sum_{i\in I_k}n_i(t),\ k=1,\dots,r$, then the random process $(N_1(t),\dots,N_r(t))$ has limit distribution $\prod_{k=1}^{r}P_k(N_k)$ where*

$$P_k(N_k)=\mathbf{e}^{(}-R_k)\frac{R_k^{N_k}}{N_k!},\ R_k=\sum_{i\in I_k}\rho(i), \tag{5}$$

and almost surely

$$\lim_{T\to\infty}\frac{\int_0^T N_j(t)dt}{T}=R(j),\ j=1,\dots,r. \tag{6}$$

Proof. The formula (5) arises from the formula (3) and from well known fact that a sum of independent random variables x,y with Poisson distributions with the parameters a,b is a random variable with Poisson distribution with the parameter $a+b$. Indeed assume that generating functions of random variables x,y are $\varphi_x(z)=\mathbf{e}^{(}a(z-1)),\ \varphi_y(z)=\mathbf{e}^{(}b(z-1))$ then $\varphi_{x+y}(z)=\varphi_x(z)\varphi_y(z)=\mathbf{e}^{(}(a+b)(z-1))$. The formula (6) arises from (4), (5). □

Remark. Denote $N(t)=\sum_{i\in I}n_i(t)$, then from Theorem 1 the discrete random process $N(t)$ has Poisson limit distribution with the parameter $R=\sum_{i\in I}\rho(i)$. Assume that k-the input customer remains the time η_k in the network with the mean $f=M\eta_k$. The quantity f may be interpreted as mean life time of some individual. So from [5, § 31, Theorem 6] we have $R=\lambda(0)f$. Consequently if there is statistical estimate of input Poisson flow intensity $\lambda(0)$ then using Theorem 2 it is possible to estimate the parameter R and so them mean individual life time f. This statement may be spread onto nodes subset $I'\subset I$ where $\theta(i,j)=0,\ i\in I\setminus I',\ j\in I'$ and η_k is time interval of k-the input customer stay in I'.

Example. Analogously with [2] assume that the route matrix Θ contains the following nonzero elements:

$$\theta(1,2), \theta(1,0) = 1 - \theta(1,2);\ \theta(2,3), \theta(2,0) = 1 - \theta(2,3); \ldots;$$

$$\theta(m-1,m), \theta(m-1,0) = 1 - \theta(m-1,m);\ \theta(m,0) = 1.$$

Then it is easy to obtain that

$$\lambda(1) = \lambda,\ \lambda(2) = \lambda(1)\theta(1,2),\ \lambda(3) = \lambda(2)\theta(2,3), \ldots, \lambda(m) = \lambda(m-1)\theta(m-1,m),$$

$$\rho(i) = \frac{\lambda(i)}{\mu(i)},\ i = 1, \ldots, m,$$

with the mean life time

$$M = \frac{1}{\mu(1)} + \frac{\theta(1,2)}{\mu(2)} + \frac{\theta(1,2)\theta(2,3)}{\mu(3)} + \ldots + \frac{\prod_{k=1}^{m-1} \theta(k,k+1)}{\mu(m)}.$$

3. CLOSED NETWORKS

Consider now closed queueing network with the nodes set S, the route matrix Θ, n customers circulating in the network and n (that is equivalent infinity) servers in each node with service intensity μ_i on a server of i-th node, $i \in S$. Assume that for fixed $\lambda(0) > 0$ positive numbers $\lambda(1), \ldots, \lambda(m)$ create single solution $(\lambda(1), \ldots, \lambda(m))$ of the system (2). Denote $\rho(0) = \lambda(0)/\mu(0)$, then discrete Markov process $(n_0(t), \ldots, n_m(t))$, describing number of customers in nodes of the closed network has polynomial limit distribution [4, Example 1.29]

$$P(n_0, \ldots, n_m) = n! \prod_{i=0}^{m} \frac{d_i^{n_i}}{n_i!},\ d_i = \frac{\rho(i)}{\rho(0) + \ldots + \rho(m)},\ i \in S,\ \sum_{i=0}^{m} n_i = n. \tag{7}$$

Lemma. *The random process* $(n_0(t) + n_1(t), n_2(t), n_3((t), \ldots, n_m(t))$ *has limit distribution*

$$P(n_0 + n_1, n_2, n_3, \ldots, n_m) = n! \frac{(d_0 + d_1)^{n_0+n_1}}{(n_0 + n_1)!} \prod_{i=2}^{m} \frac{d_i^{n_i}}{n_i!}. \tag{8}$$

Proof. Lemma statement is based on binomial theorem. □

Theorem 3. *Almost surely*

$$\lim_{T\to\infty} \frac{\int_0^T n_i(t)dt}{T} = nd_i,\ i \in I_0. \tag{9}$$

Proof. Using the formula (8) by mathematical induction it is easy to prove that the random process $(n_0(t), n_1(t) + \ldots + n_m(t))$ has limit distribution

$$P(n_0, n_1 + \ldots + n_m) = \frac{n! d_0^{n_0} (d_1 + \ldots + d_m)^{n-n_0}}{n_0!(n-n_0)!}.$$

Consequently the random process $n_0(t)$ has limit Bernoulli distribution with the mean nd_0. Analogously it is possible to prove that the random process $n_i(t)$ has limit Bernoulli distribution with the mean $nd_i,\ i = 1, \ldots, m$. So from the law of large numbers for ergodic discrete Markov process [4, Theorem 1.2] we obtain the formula (9). □

Theorem 4. *Suppose that $I_1, \ldots, I_r$ create a decomposition of the set S onto non-intersected subsets. Denote $N_k(t) = \sum_{i \in I_k} n_i(t),\ k = 1, \ldots, r$, then the random process $(N_1(t), \ldots, N_r(t))$ has polynomial limit distribution*

$$P(N_1, \ldots, N_r) = n! \prod_{k=1}^{r} \frac{D_k^{N_k}}{N_k!},\ D_k = \sum_{i \in I_k} d_i,\ k = 1, \ldots, r, \tag{10}$$

and almost surely

$$\lim_{T \to \infty} \frac{\int_0^T N_k(t) dt}{T} = nD_k,\ k = 1, \ldots, r. \tag{11}$$

Proof. The formula (10) may be obtained from the formula (8) by mathematical induction. The formula (11) arises from the formula (10) and Theorem 3. □

4. CONCLUSION

Theorems 1 - 4 allow to aggregate nodes of initial network and so to simplify statistical estimate of its parameters. As a result random process, which describes customers motion along nodes-states and which is defined by the route matrix Θ with the dimension $m \times m$ and by service intensities $\mu(i),\ i = 1, \ldots, m$, is replaced by statistical model of customers distribution in aggregated nodes. It allows not only to get rid of complicated solution of the system (2) but to simplify a procedure of observation and a collection of necessary information. It is well known that to define a state i is more difficult than to define its belonging to the subset $I_j \subseteq I$. More over it is possible to construct sufficiently simple estimate of mean life time also. This quantity is important in analysis of efficiency indexes.

REFERENCES

1. *Liu Y.W. , Kapur K.C.* New Models and Measures for Reliability of Multi-state Systems. In Handbook of Performability Engineering. Krishna B. Misra (ed). Springer Verlag. 2008. P. 431-445.

2. *Han I.S., Gulevich A.P.* Optimization of strategies of noninfectious diseases preventionism// Bulletin of Pacific State Economical University. 2008. No. 2. P. 10-22. (In Russian).
3. *Basharin G.P., Tolmachev A.L.* Queueing Network Theory and Its Applications to the Analysis of Information-Computing Networks. In Itogi Nauki Tekh. Ser. Teor. Veroyatn. Mat. Stat. Teor. Kibern. Vol. 21. Moscow: VINITI. 1983. P. 3-119. (In Russian).
4. *Serfozo R.* Introduction to Stochastic Networks. New York: Springer Verlag. 1999.
5. *Borovkov A.A.* Probability processes in queueing theory. Moscow: Nauka. 1972. (In Russian).

ГИСТЕРЕЗИСНАЯ СТРАТЕГИЯ ДЛЯ СИСТЕМЫ С ПОВТОРНЫМИ ВЫЗОВАМИ

И. Усар*, **И. Макушенко**

Киевский национальный университет имени Тараса Шевченко
Киев, Украина
* usar@unicyb.kiev.ua

Рассматривается марковская система с повторными вызовами и управляемой интенсивностью входного потока. В явном виде через параметры модели найдены стационарные вероятности и указаны условия существования стационарного режима.

Ключевые слова: стохастическая система, повторные требования, гистерезисная стратегия, стационарный режим.

1. ВВЕДЕНИЕ

Результаты для систем с повторными вызовами составляют один из важных разделов теории массового обслуживания. Основные из этих результатов систематизированы в монографии [1]. Математические модели систем с повторными вызовами имеют широкое применение в экономике, транспорте, в практике проектирования компьютерных сетей (локальных и глобальных), системах с заказом в телефонии и мобильной связи, для описания процесса посадки воздушных судов и др. (См., например, [1, 2, 3]).

Характерная особенность рассматриваемых систем заключается в том, что требования, поступившие в занятую систему, через случайный промежуток времени повторяют попытку получить обслуживание. Такие требования становятся источниками повторных вызовов, генерирующих вторичный поток вызовов.

2. ОПИСАНИЕ СИСТЕМЫ

Рассматривается система обслуживания типа $M_Q/M/m/\infty$ с повторными вызовами, состоящая из m идентичных обслуживающих приборов, работающих независимо. Время обслуживания имеет показательное распределение с параметром μ. Интенсивность λ_j потока первичных вызовов зависит от j - колличества источников повторных вызовов. Каждый повторный вызов через показательное время с параметром ν повторяет попытку получить обслуживание. Кроме того, поток первичных вызовов подчиняется гистерезисной стратегии, которая может быть описана следующим образом. Пусть $h_1 \leq h_2$ будут фиксированные натуральные числа (пороги). Если колличество повторных вызовов $j \leq h_1 - 1$, то интенсивность

входного потока равна λ_1 и система работает в первом режиме. Если $j \geq h_2$, то интенсивность входного потока равна λ_2 и система работает во втором режиме. А если $h_1 \leq j \leq h_2 - 1$, то система сохраняет тот режим, в котором она работала в предыдущий момент времени. Частным случаем гистерезисных стратегий является класс пороговых стратегий (для них $h_1 = h_2$). Такие системы рассматривались в работе [4].

Описанную модель можно изобразить трехмерным процессом Маркова $Q(t) = (Q_1(t), Q_2(t), Q_3(t))$ с непрерывным временем в фазовом пространстве $S = \{0, 1, ..., m\} \times \{0, 1, ...\} \times \{1, 2\}$, где $Q_1(t)$ - число занятых приборов в момент времени t, $Q_2(t)$ - число источников повторных требований, $Q_3(t)$ - режим работы системы в момент времени t. Если $Q_3(t) = 1$, то система работает в первом режиме, а если $Q_3(t) = 2$, то система работает во втором режиме.

3. УСЛОВИЯ СУЩЕСТВОВАНИЯ И ФОРМУЛЫ ДЛЯ ЭРГОДИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Выясним условия существования стационарного режима для процесса $Q(t)$.

Лемма 1. *Пусть* $\lambda_2 = \limsup_j \lambda_j < \infty$. *Тогда при* $\lambda_2/m\mu < 1$ *существует эргодическое распределение для процесса* $Q(t)$ *и оно совпадает с единственным стационарным.*

Сначала найдем эргодическое распределение для системы, в которой колличество повторных требований ограничено некоторым числом N, а затем, перейдя к пределу $N \to \infty$, получим это распределение без ограничений на длину очереди. Для $M_Q/M/m/N$ эргодическое распределение существует и будем обозначать его $\pi_{ij}^{(r)}(N) = \lim_{t\to\infty} \mathbf{P}\{Q_1(t) = i, Q_2(t) = j, Q_3(t) = r\}$.

Введем следующие обозначения:

$$e_i(m) = (\delta_{i0}, \delta_{i1}, ..., \delta_{im-1})^T, \quad \delta_{ij} = \begin{cases} 1, & \text{при} \quad i = j, \\ 0, & \text{при} \quad i \neq j, \end{cases} \qquad e(m) = (1, 1, ..., 1)^T,$$

$$\pi_j^{(r)} = \left(\pi_{0j}^{(r)}, \pi_{1j}^{(r)}, ..., \pi_{m-1j}^{(r)}\right)^T.$$

Пусть $A_{rj} = \| a_{ik}^j(r) \|_{i,k=0}^{m-1}$, $j = 0, 1, \ldots, N-1$ - тридиагональная матрица вида

$$a_{ik}^j(r) = \begin{cases} \lambda^{(r)} + i\mu + j\nu, & \text{при} \quad k = i, i = 0, 1, ..., m-1, \\ -\lambda^{(r)}, & \text{при} \quad k = i-1, i = 0, 1, ..., m-1, \\ -(i+1)\mu, & \text{при} \quad k = i+1, i = 1, 2, ..., m-2, \\ 0, & \text{в остальных случаях,} \end{cases}$$

$$B_{rj} = \left\| b_{ik}^{j}(r) \right\|_{i,k=0}^{m-1}, \quad b_{ik}^{j}(r) = \begin{cases} (j+1)\nu, & \text{при} \quad k = i-1, i \neq m-1, \\ \dfrac{(j+1)\nu m\mu}{\lambda^{(r)}}, & \text{при} \quad k \neq m-2, i = m-1, \\ \dfrac{(j+1)\nu(\lambda^{(r)} + m\mu)}{\lambda^{(r)}}, & \text{при} \quad k = m-2, i = m-1, \\ 0, & \text{в остальных случаях,} \end{cases}$$

$$C_{rj} = \left\| c_{ik}^{j}(r) \right\|_{i,k=0}^{m-1}, \qquad c_{ik}^{j}(r) = \begin{cases} \dfrac{h_1 \nu m\mu}{\lambda^{(r)}}, & \text{при} \quad i = m-1, \\ 0, & \text{в остальных случаях,} \end{cases}$$

$$D = \| d_{ik} \|_{i,k=0}^{m-1}, \qquad d_{ik} = \begin{cases} 1, & \text{при} \quad k = 0, \ i = 0, \\ a_{i-1k}^{N}(2), & \text{в остальных случаях.} \end{cases}$$

$$\Delta_{1j}^{(1)} = \left(\prod_{i=j}^{h_1-1} A_{1i}^{-1} B_{1i} \right) \left(\left[\left(\prod_{i=h_1}^{h_2-2} A_{1i}^{-1} B_{1i} \right) A_{1h_2-1}^{-1} C_{1h_2-1} + \sum_{k=h_1}^{h_2-2} \left(\prod_{i=h_1}^{k-1} A_{1i}^{-1} B_{1i} \right) A_{1k}^{-1} C_{1k} \right] + \right.$$
$$\left. + E \right) \left[E + \sum_{k=h_1}^{h_2-1} \left(\prod_{i=h_1}^{k-1} A_{2i}^{-1} B_{2i} \right) A_{2k}^{-1} C_{2k} \right]^{-1} \left(\prod_{i=h_1}^{N-1} A_{2i}^{-1} B_{2i} \right) D^{-1} e_0(m), \ j = 0, 1, ..., h_1 - 1,$$

$$\Delta_{2j}^{(1)} = \left[\left(\prod_{i=j}^{h_2-2} A_{1i}^{-1} B_{1i} \right) A_{1h_2-1}^{-1} C_{1h_2-1} + \sum_{k=j}^{h_2-2} \left(\prod_{i=j}^{k-1} A_{1i}^{-1} B_{1i} \right) A_{1k}^{-1} C_{1k} \right] \times$$
$$\times \left[E + \sum_{k=h_1}^{h_2-1} \left(\prod_{i=h_1}^{k-1} A_{2i}^{-1} B_{2i} \right) A_{2k}^{-1} C_{2k} \right]^{-1} \left(\prod_{i=h_1}^{N-1} A_{2i}^{-1} B_{2i} \right) D^{-1} e_0(m), \ j = h_1, ..., h_2 - 1,$$

$$\Delta_{1j}^{(2)} = \left(E - \left[\sum_{k=j}^{h_2-1} \left(\prod_{i=j}^{k-1} A_{2i}^{-1} B_{2i} \right) A_{2k}^{-1} C_{2k} \right] \left[E + \sum_{k=h_1}^{h_2-1} \left(\prod_{i=h_1}^{k-1} A_{2i}^{-1} B_{2i} \right) A_{2k}^{-1} C_{2k} \right]^{-1} \right) \times$$
$$\times \left(\prod_{i=j}^{N-1} A_{2i}^{-1} B_{2i} \right) D^{-1} e_0(m), \quad j = h_1, ..., h_2 - 1,$$

$$\Delta_{2j}^{(2)} = \left(\prod_{i=j}^{N-1} A_{2i}^{-1} B_{2i} \right) D^{-1} e_0(m), \quad j = h_2, ..., N-1,$$

$$\Delta_{j}^{(1)} = \begin{cases} \Delta_{1j}^{(1)}, & \text{при} \quad j = 0, ..., h_1 - 1, \\ \Delta_{2j}^{(1)} & \text{при} \quad j = h_1, ..., h_2 - 1, \end{cases} \qquad \Delta_{j}^{(2)} = \begin{cases} \Delta_{1j}^{(1)}, & \text{при} \quad j = h_1, ..., h_2 - 1, \\ \Delta_{2j}^{(2)} & \text{при} \quad j = h_2, ..., N-1. \end{cases}$$

Проверим условия существования обратных матриц, которые используются при определении векторов $\Delta_j^{(1)}$ и $\Delta_j^{(2)}$.

Лемма 2. *Матрицы* $A_{rj},\ r=1,2,\ j=0,1,...,N$ *и* D *невырожденные.*

Найдем теперь явные формулы для эргодического распределения процесса $Q(t),\ t\geq 0$, при условии, что оно существует.

Теорема 1. *Стационарные вероятности системы имеют вид*

$$\pi_j^{(1)}(N)=\Delta_j^{(1)}\pi_{0N}^{(2)}(N),\ j=0,...h_2-1,\ \pi_j^{(2)}(N)=\Delta_j^{(2)}\pi_{0N}^{(2)}(N),\ j=h_1,...N-1,$$

$$\pi_N^{(2)}(N)=D^{-1}e_0(m)\pi_{0N}^{(2)}(N),$$

$$\pi_{mj}^{(1)}(N)=\frac{\nu}{\lambda^{(1)}}\left((j+1)\Delta_{j+1}^{(1)}+\alpha_1(j)h_1\Delta_{h_1}^{(2)}\right)e(m)^T\pi_{0N}^{(2)}(N),\ j=0,...,h_2-1,$$

$$\pi_{mj}^{(2)}(N)=\frac{\nu}{\lambda^{(2)}}\left((j+1)\Delta_{j+1}^{(2)}-\alpha_2(j)h_1\Delta_{h_1}^{(2)}\right)e(m)^T\pi_{0N}^{(2)}(N),\ j=h_1,...,N-1,$$

$$\pi_{mN}^{(2)}(N)=\frac{\left(\lambda^{(2)}+N\nu+(m-1)\mu\right)e_{m-1}(m)^T-\lambda^{(2)}e_{m-2}(m)^T}{m\mu}D^{-1}e_0(m)\pi_{0N}^{(2)}(N),$$

где

$$\pi_{0N}^{(2)}(N)=\left[\sum_{j=0}^{h_2-1}\frac{\lambda^{(1)}\Delta_j^{(1)}+\nu\left((j+1)\Delta_{j+1}^{(1)}+\alpha_1(j)h_1\Delta_{h_1}^{(2)}\right)}{\lambda^{(1)}}e(m)^T+\right.$$

$$+\sum_{j=h_1}^{N-1}\frac{\lambda^{(2)}\Delta_j^{(2)}+\nu\left((j+1)\Delta_{j+1}^{(2)}-\alpha_2(j)h_1\Delta_{h_1}^{(2)}\right)}{\lambda^{(2)}}e(m)^T+$$

$$\left.+\frac{e(m)^Tm\mu+\left(\lambda^{(2)}+N\nu+(m-1)\mu\right)e_{m-1}(m)^T-\lambda^{(2)}e_{m-2}(m)^T}{m\mu}D^{-1}e_0(m)\right]^{-1},$$

$$\alpha_1(j)=\begin{cases}1, & j>h_1-2,\\ 0, & j\leq h_1-2,\end{cases}\qquad \alpha_2(j)=\begin{cases}1, & j<h_2,\\ 0, & j\geq h_2.\end{cases}$$

При выполнении условий леммы 1 и $N\to\infty$ стационарные вероятности $\pi_{ij}^{(r)}(N)$ приближают соответствующие вероятности для системы $M_Q/M/m/\infty$. Таким образом, теорема содержит эффективный алгоритм рекурентного типа для подсчета стационарных характеристик.

Рассмотрим применение теоремы для решения оптимизационной задачи.

Пусть $S_1(t,h_1,h_2)$ - число вызовов, облуживание которых завершено в системе за время t, $S_2(t,h_1,h_2)$ - число вызовов, которые получили отказ в обслуживании и стали повторными вызовами, $S_3(t,h_1,h_2)$ - число переключений интенсивности входного потока.

Если существуют пределы $\lim\limits_{t\to\infty}t^{-1}S_i(t,h_1,h_2)$, то будем обозначать их $S_i\left(h_1,h_2\right),\ i=1,2,3.$

Рассмотрим оптимизационную задачу:

$$W\left(h_1,h_2\right)=C_1S_1\left(h_1,h_2\right)-C_2S_2\left(h_1,h_2\right)-C_3S_3\left(h_1,h_2\right)\to\max,$$

где C_1 - прибыль полученная от обслуживания одного требования, C_2 - штраф за отказ в обслуживании, C_3 - штраф за переключение интенсивности входного потока.

Решением задачи есть такие пороги h_1 и h_2, которые максимизируют среднюю прибыль от работы системы. Подобные оптимизационные задачи рассматривалсь в работах [5, 6].

4. ЗАКЛЮЧЕНИЕ

В заключение хотелось бы отметить, что полученные результаты имеют большое значение для практического применения. На основании найденных формул для стационарных вероятностей можно построить эффективные алгоритмы для решения оптимизационных задач в классе гистерезисных стратегий.

ЛИТЕРАТУРА

1. *Falin G. I., Templeton J. G. C.* Retrial queues. London: Chapman & Hall, 1997.
2. *Уолрэнд Дж.* Введение в теорию массового обслуживания. Москва: Мир, 1993.
3. *Anisimov V. V., Artalejo J. R.* Analysis of Markov multiserver retrial queues with negative arrivals // Queuing Systems. 2001. № 39. С. 157–182.
4. *Лебедев Е. А., Усар И. Я.* Про системы с повторными требованиями и управляемыми входящими потоками // Доклады НАН Украины. 2009. № 5. С. 52–59.
5. *Клименок В. И.* Оптимизация динамического управления режимом работы информационно-вычислительных систем с повторными вызовами // Автоматика и вычислит. техника. 1990. № 1. С. 25–30.
6. *Дудин А. Н., Клименок В. И.* Оптимизация динамического управления входной нагрузкой в узле информационно-вычислительной сети // Автоматика и вычислит. техника. 1991. № 2. С. 25–31.

СИСТЕМА $BMAP/G/1$ СО ШЛЮЗОВЫМ ОБСЛУЖИВАНИЕМ И АДАПТИВНЫМИ ОТДЫХАМИ

В. Вишневский[1], А. Дудин[2,⋆], В. Клименок[2], О. Семенова[1], Ю. Синюгина[3]

[1] *ЗАО «Информационные и сетевые технологии»,*
[2] *Белорусский государственный университет,*
[3] *Гомельский государственный университет им. Ф. Скорины*
[1] *Москва, Россия*
[2] *Минск, Республика Беларусь*
[3] *Гомель, Республика Беларусь*
[⋆] dudin@bsu.by

В работе рассмотрена система $BMAP/G/1$ с отдыхами прибора и шлюзовым обслуживанием. Длительность отдыха прибора зависит от того, сколько раз подряд система оказывалась пустой в предыдущие моменты окончания отдыха. Получены стационарные распределения: числа заявок в системе в момент завершения отдыха, состояний системы в произвольный момент времени, времени ожидания.

Данная работа была частично поддержана Российским фондом фундаментальных исследований (грант № 10-07-90006) и Белорусским республиканским фондом фундаментальных исследований (грант № Ф10Р-010).

Ключевые слова: групповой марковский входной поток, адаптивные отдыхи, шлюзовое обслуживание, матрично-аналитические методы.

1. ВВЕДЕНИЕ

В связи с интенсивным развитием широкополосных беспроводных сетей передачи информации в последние годы значительно возрос интерес к моделям систем массового обслуживания с несколькими очередями и общим обслуживающим прибором (системам поллинга). Отдельную очередь системы поллинга можно рассматривать как систему массового обслуживания (СМО) с отдыхами прибора, где под отдыхом понимается время, в течение которого прибор обслуживает другие очереди системы. С точки зрения практики значительный интерес представляют модели с так называемым адаптивным опросом, в которых длительность отдыха прибора зависит от того, были ли в данной очереди заявки на обслуживание в момент завершения предыдущего отдыха. Исследованию таких моделей посвящена данная работа.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассмотрим однолинейную СМО с двумя бесконечными буферами, соединенными шлюзом. На вход системы поступает $BMAP$-поток заявок, характеризующийся управляющим процессом $\nu_t, t \geq 0$, с пространством состояний $\{0, 1, \ldots, W\}$ и матричной производящей функцией $D(z) = \sum\limits_{k=0}^{\infty} D_k z^k$, $|z| < 1$. Матрицы D_k задают интенсивности переходов процесса ν_t, которые сопровождаются приходом группы заявок размера k, $k \geq 0$. Матрица $D(1)$ является генератором процесса $\nu_t, t \geq 0$.

Средняя интенсивность $BMAP$-потока λ задается формулой $\lambda = \boldsymbol{\theta} D'(1)\mathbf{e}$, где $\boldsymbol{\theta}$ является единственным решением системы линейных алгебраических уравнений $\boldsymbol{\theta} D(1) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$, $\mathbf{e}$ - вектор-столбец соответствующей размерности, состоящий из единиц, $\mathbf{0}$ - вектор-строка соответствующей размерности, состоящий из нулей.

Поступая в систему, группа заявок попадает в первый буфер (буфер-1). Обслуживание заявок из этой группы будет возможно только после открытия шлюза между буферами, когда все заявки, находящиеся в буфере-1, мгновенно переместятся в буфер-2 и шлюз закроется. В момент открытия шлюза возможны две ситуации. Первая предполагает, что буфер-1 окажется пустым и в предыдущие r-1 моменты открытия шлюза он также был пуст, $r \geq 1$. В этом случае прибор уходит на отдых, называемый отдыхом типа-r. Его длительность характеризуется функцией распределения (ФР) $H_r(t)$ с преобразованием Лапласа-Стилтьеса (ПЛС) $h_r(s) = \int\limits_0^{\infty} e^{-st} dH_r(t)$, $Re\ s > 0$, и конечными начальными моментами $h_k^{(r)} = \int\limits_0^{\infty} t^k dH_r(t)$. Если же буфер-1 не был пуст в момент открытия шлюза, то заявки, находящиеся в нем, перемещаются в буфер-2 и начинается обслуживание прибором заявок из буфера-2 в порядке их поступления в систему. Время обслуживания заявок имеет ФР $B(t)$ с ПЛС $\beta(s) = \int\limits_0^{\infty} e^{-st} dB(t)$, $Re\ s > 0$, и начальными конечными моментами $b_k = \int\limits_0^{\infty} t^k dB(t)$. Обслуживание заявок продолжается до тех пор, пока буфер-2 не окажется пустым. После этого прибор уходит на отдых, называемый *обычным* или отдыхом типа-0, длительность которого характеризуется ФР $H_0(t)$ с ПЛС $h_0(s) = \int\limits_0^{\infty} e^{-st} dH_0(t)$, $Re\ s > 0$, и начальными конечными моментами $h_k^{(0)} = \int\limits_0^{\infty} t^k dH_0(t)$. После окончания отдыха шлюз между буферами вновь открывается.

Проанализируем описанную систему. Будем рассматривать процесс

$$\zeta(t) = \{i_1(t), i_2(t), r(t), \nu_t\}, \quad t \geq 0,$$

где $i_k(t)$ – число заявок в буфере-k, а процесс $r(t)$ характеризует состояние прибора:

$$r(t) = \begin{cases} *, & \text{если в момент } t \text{ прибор занят обслуживанием заявки,} \\ r, & \text{если в момент } t \text{ прибор находится на отдыхе типа-}r,\ r \geq 0. \end{cases}$$

Понятно, что $i_2(t) = 0$, если $r(t) \neq *$. Процесс $\zeta(t)$, $t \geq 0$, не является марковским, и для его исследования мы применим метод вложенных цепей.

3. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВЛОЖЕННОЙ ЦЕПИ МАРКОВА

Пусть t_n есть n-й момент окончания отдыха, j_n есть число заявок в буфере-1 в момент $t_n - 0$ и r_n принимает значение 0, если буфер-1 не был пуст в момент t_{n-1}, или значение r, если буфер-1 был пуст в моменты t_{n-r}, t_{n-r+1}, ... , t_{n-1}, $r \geq 1$, и $\nu_n = \nu_{t_n}$.

Процесс $\xi_n = \{j_n, r_n, \nu_n\}$, $n \geq 1$, является цепью Маркова. Пусть матрицы $P_{(i,r),(j,r')}$, $i, j \geq 0$, $r, r' \geq 0$, составлены из ее одношаговых переходных вероятностей

$$P_{(i,r),(j,r')} = P\{j_{n+1} = j; r_{n+1} = r'; \nu_{n+1} = \nu' | j_n = i; r_n = r; \nu_n = \nu\}.$$

Ненулевые матрицы $P_{(i,r),(j,r')}$ имеют вид:

$$P_{(i,r),(j,0)} = \int_0^\infty P(j,t) dB^{*(i)} * H_0(t), \ i \geq 1, j \geq 0, r \geq 0, \tag{1}$$

$$P_{(0,r),(j,r+1)} = \int_0^\infty P(j,t) dH_{r+1}(t), \ j \geq 0, r \geq 0, \tag{2}$$

где матрицы $P(l,t)$ определяются как коэффициенты разложения $e^{D(z)t} = \sum\limits_{l=0}^{\infty} P(l,t) z^l$, $B^{*(i)}(t)$ есть свертка i-го порядка ФР $B(t)$ и $A * H(t)$ есть свертка ФР $A(t)$ и $H(t)$.

При выполнении условия $\rho = \lambda b_1 < 1$ существуют стационарные вероятности состояний цепи

$$q(j,r,\nu) = \lim_{n\to\infty} P\{j_n = j, \ r_n = r, \ \nu_n = \nu\}, \ j \geq 0, r \geq 0, \nu = \overline{0, W}.$$

Сгруппируем эти вероятности в векторы-строки $\mathbf{q}(j,r) = (q(j,r,0), \ \ldots, \ q(j,r,W))$. Эти векторы удовлетворяют системе линейных алгебраических уравнений

$$\mathbf{q}(j,0) = \sum_{r=0}^{\infty} \sum_{i=1}^{\infty} \mathbf{q}(i,r) \int_0^\infty P(j,t) dB^{*(i)} * H_0(t), \ j \geq 0, \tag{3}$$

$$\mathbf{q}(j,r+1) = \mathbf{q}(0,r) \int_0^\infty P(j,t) dH_{r+1}(t), j \geq 0, r \geq 0. \tag{4}$$

Умножая уравнения системы (3), (4) на соответствующие степени z и суммируя, можно убедиться, что векторные производящие функции $\mathbf{q}_r(z) = \sum\limits_{j=0}^{\infty} \mathbf{q}(j,r) z^j$, $|z| <$

$1, r \geq 0$ удовлетворяют следующей системе функциональных уравнений:

$$\mathbf{q}_0(z) = \sum_{r=0}^{\infty} (\mathbf{q}_r(\beta(-D(z))) - \mathbf{q}_r(0))\, h_0(-D(z)), \tag{5}$$

$$\mathbf{q}_{r+1}(z) = \mathbf{q}_r(0) h_{r+1}(-D(z)), r \geq 0, \tag{6}$$

где $\beta(-D(z)) = \int\limits_0^{\infty} e^{D(z)t} dB(t)$, $h_r(-D(z)) = \int\limits_0^{\infty} e^{D(z)t} dH_r(t), r \geq 0$.

Система уравнений (5), (6) может быть переписана в следующем виде:

$$\mathbf{q}_0(z) = \mathbf{q}_0(\beta(-D(z)))h_0(-D(z)) + \mathbf{q}_0(0)A(z), \tag{7}$$

$$\mathbf{q}_{r+1}(z) = \mathbf{q}_0(0)\Theta_r h_{r+1}(-D(z)), r \geq 0. \tag{8}$$

где

$$A(z) = h_0(-D(z)) \left[\sum_{r=0}^{\infty} \Theta_r (h_{r+1}(-D(\beta(-D(z)))) - h_{r+1}(-D(0))) - I \right],$$

$$\Theta_r = \prod_{j=1}^{r} h_j(-D(0)).$$

Для решения матричного функционального уравнения (7), введем в рассмотрение последовательность матриц $N_m(z)$, задаваемых рекурсией

$$N_0(z) = zI, \;\; N_{m+1}(z) = \beta(-D(N_m(z))), \; m \geq 0. \tag{9}$$

В [3] показано, что при любом z, $|z| \leq 1$, последовательность матриц $N_m(z)$ сходится к матрице G, определяемой как решение матричного уравнения

$$G = \beta(-D(G)) = \int\limits_0^{\infty} e^{D(G)t} dB(t).$$

Последовательно подставляя матрицы $N_m(z)$ вместо скалярного аргумента z в функциональное уравнение (7), мы получаем соотношения

$$\mathbf{q}_0(N_0(z)) = \mathbf{q}_0(N_n(z)) \prod_{m=0}^{n-1} h_0(-D(N_m(z)) + \mathbf{q}_0(0) \sum_{m=0}^{n-1} A(N_m(z)) \prod_{k=0}^{m-1} h_0(-D(N_k(z))).$$

Устремляя параметр n к бесконечности, получаем выражение

$$\mathbf{q}_0(z) = \mathbf{q}_0(G) \prod_{l=0}^{\infty} h_0(-D(N_l(z))) + \mathbf{q}_0(0) \sum_{r=0}^{\infty} A(N_r(z)) \prod_{m=0}^{r-1} h_0(-D(N_m(z))). \tag{10}$$

Для вычисления векторов $\mathbf{q}_0(0)$ и $\mathbf{q}_0(G)$ подставим $z = G$ в уравнении (7) и $z = 0$ в (10). В результате мы получим следующее уравнение:

$$(\mathbf{q}_0(0), \mathbf{q}_0(G)) = (\mathbf{q}_0(0), \mathbf{q}_0(G)) \begin{pmatrix} \sum\limits_{r=0}^{\infty} A(N_r(0)) \prod\limits_{m=0}^{r-1} h_0(-D(N_m(0))) & A(G) \\ \prod\limits_{l=0}^{\infty} h_0(-D(N_l(0))) & h_0(-D(G)) \end{pmatrix}. \tag{11}$$

Еще одно уравнение для компонент неизвестного вектора $(\mathbf{q}_0(0), \mathbf{q}_0(G))$ мы получаем из условия нормировки

$$\mathbf{q}_0(0)\Big[\sum_{r=0}^{\infty} A(N_r(1)) \prod_{m=0}^{r-1} h_0(-D(N_m(1))) + \sum_{r=0}^{\infty} \Theta_r h_{r+1}(-D(1))\Big]\mathbf{e}+$$

$$+\mathbf{q}_0(G) \prod_{l=0}^{\infty} h_0(-D(N_l(1)))\mathbf{e} = 1. \tag{12}$$

Теорема 1. *Векторная производящая функция $\mathbf{q}_0(z)$ задается формулой (10), где векторы $\mathbf{q}_0(0), \mathbf{q}_0(G)$ определяются как решение системы (11)-(12).*

Теорема 2. *Векторная производящая функция $\mathbf{q}(z) = \sum\limits_{r=0}^{\infty} \mathbf{q}_r(z)$ задается формулой*

$$\mathbf{q}(z) = \mathbf{q}_0(G) \prod_{l=0}^{\infty} h_0(-D(N_l(z)))+$$

$$+\mathbf{q}_0(0) \sum_{r=0}^{\infty} \Big[A(N_r(z)) \prod_{m=0}^{r-1} h_0(-D(N_m(z))) + \Theta_r h_{r+1}(-D(z))\Big]. \tag{13}$$

4. РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ СОСТОЯНИЙ СИСТЕМЫ В ПРОИЗВОЛЬНЫЙ МОМЕНТ ВРЕМЕНИ

Обозначим

$$p(i_1, i_2, *, \nu) = \lim_{t\to\infty} P\{i_1(t) = i_1, i_2(t) = i_2, r_t = *, \nu_t = \nu\}, \quad i_2 \geq 1,$$

$$p(i_1, r, \nu) = \lim_{t\to\infty} P\{i_1(t) = i_1, r_t = r, \nu_t = \nu\}, \ i_1 \geq 0, \ r \geq 0, \nu = \overline{0, W}.$$

Введем векторы

$$\mathbf{p}(i_1, i_2, *) = (p(i_1, i_2, *, 0), \ldots, p(i_1, i_2, *, W)), \quad \mathbf{p}(i_1, r) = (p(i_1, r, 0), \ldots, p(i_1, r, W)).$$

Теорема 3. *Векторы стационарных вероятностей $\mathbf{p}(i_1, i_2, *)$, $\mathbf{p}(i_1, r)$, $r \geq 0$ выражаются через векторы $\mathbf{q}(i, r)$, $i \geq 0$, следующим образом:*

$$\mathbf{p}(i_1, i_2, *) = \tau^{-1}\Big[\sum_{j=i_2+1}^{\infty} \sum_{r=0}^{\infty} \mathbf{q}(j, r) \sum_{l=0}^{i_1} \int_0^{\infty} P(l, t) dB^{*(j-i_2)}(t) \int_0^{\infty} P(i_1 - l, t)(1 - B(t))dt+$$

$$+\sum_{r=0}^{\infty}\mathbf{q}(i_2,r)\int\limits_0^{\infty}P(i_1,t)(1-B(t))dt\Big],$$

$$\mathbf{p}(i_1,0)=\tau^{-1}\sum_{j=1}^{\infty}\sum_{r=0}^{\infty}\mathbf{q}(j,r)\sum_{l=0}^{i_1}\int\limits_0^{\infty}P(l,t)dB^{*(j)}(t)\int\limits_0^{\infty}P(i_1-l,t)(1-H_0(t))dt,$$

$$\mathbf{p}(i_1,r)=\tau^{-1}\mathbf{q}(0,r-1)\int\limits_0^{\infty}P(i_1,t)(1-H_r(t))dt,$$

где среднее время τ между моментами окончания отдыхов вычисляется по формуле

$$\tau=b_1\mathbf{q}'(z)|_{z=1}\mathbf{e}+h_1^{(0)}(1-\mathbf{q}(0))\mathbf{e}+\mathbf{q}_0(0)\sum_{r=0}^{\infty}\Theta_r h_1^{(r+1)}\mathbf{e}.$$

Доказательство проводится с помощью теории процессов восстановления с зависимыми циклами регенерации, см., например, [1].

Введем векторные производящие функции

$$\Pi(z,y,*)=\sum_{i_1=0}^{\infty}\sum_{i_2=1}^{\infty}\mathbf{p}(i_1,i_2,*)z^{i_1}y^{i_2},\ |z|\le 1,\ |y|\le 1,$$

$$\Pi(z,r)=\sum_{i_1=0}^{\infty}\mathbf{p}(i_1,r)z^{i_1},\ |z|\le 1,\ \ r\ge 0.$$

Теорема 4. *Векторные производящие функции $\Pi(z,y,*)$, $\Pi(z,r)$, $r\ge 0$ вычисляются по формулам*

$$\Pi(z,y,*)=\tau^{-1}(\mathbf{q}(\beta(-D(z)))-\mathbf{q}(y))y(yI-\beta(-D(z)))^{-1}(-D(z))^{-1}(I-\beta(-D(z))),$$

$$\Pi(z,0)=\tau^{-1}(\mathbf{q}(\beta(-D(z)))-\mathbf{q}(0))(-D(z))^{-1}(I-h_0(-D(z))),$$

$$\Pi(z,r)=\tau^{-1}\mathbf{q}_{r-1}(0)(-D(z))^{-1}(I-h_r(-D(z))).$$

5. РАСПРЕДЕЛЕНИЕ ВРЕМЕНИ ОЖИДАНИЯ В СИСТЕМЕ

Пусть $W(x)$ есть ФР времени ожидания произвольной заявки в системе, $w(s)=\int\limits_0^{\infty}e^{-sx}dW(x),\ Re\ s>0$ - ее ПЛС.

Теорема 5. *Преобразование Лапласа-Стилтьеса $w(s)$ времени ожидания произвольной заявки в системе задается формулой*

$$w(s)=(\lambda\tau)^{-1}(\mathbf{q}(\beta(s))(h_0(s)-1)+\sum_{r=0}^{\infty}\mathbf{q}_r(0)h_{r+1}(s)-\mathbf{q}(0)h_0(s))(1-\beta(s))^{-1}\mathcal{B}(s),\ (22)$$

где

$$\mathcal{B}(s) = (sI + D(\beta(s)))^{-1} D(\beta(s))\mathbf{e}. \tag{23}$$

Доказательство проводится с использованием техники, описанной в [2].

Среднее время ожидания W_1 произвольного запроса вычисляется по формуле

$$W_1 = \frac{1}{\lambda\tau(1-\rho)}\Bigg[\frac{\rho}{2b_1^2}\Bigg((\mathbf{q}(1)-\mathbf{q}(0))(h_2^{(0)}b_1 - h_1^{(0)}b_2) + h_1^{(0)}2b_1^2\mathbf{q}^{'}(1)+$$

$$\sum_{r=0}^{\infty}\mathbf{q}_r(0)(h_2^{(r+1)}b_1 - h_1^{(r+1)}b_2)\Bigg)+$$

$$+b_1^{-1}\Bigg(h_1^{(0)}(\mathbf{q}(1)-\mathbf{q}(0)) + \sum_{r=0}^{\infty}\mathbf{q}_r(0)h_1^{(r+1)}\Bigg)\Bigg(\tilde{I}D(1) + \hat{\mathbf{e}}\boldsymbol{\theta}(I - b_1D^{'}(1))\Bigg)^{-1}\times$$

$$\times\Bigg(\tilde{I}(\rho - b_1D^{'}(1)) + \frac{1}{2}\hat{\mathbf{e}}\boldsymbol{\theta}(b_1^2D^{''}(1) + b_2D^{'}(1))\Bigg)\Bigg]\mathbf{e}.$$

Здесь $\tilde{I}$ – диагональная матрица с диагональными элементами $\{0, 1, \ldots, 1\}$ и $\hat{\mathbf{e}}$ – вектор-столбец с элементами $\{1, 0, \ldots, 0\}$.

6. ЗАКЛЮЧЕНИЕ

В работе рассмотрена система $BMAP/G/1$ с адаптивными отдыхами прибора и шлюзовым обслуживанием. Получено стационарное распределение вероятностей состояний системы и время ожидания.

ЛИТЕРАТУРА

1. E. Cinlar, Introduction to stochastic process. N.J.: Prentice-Hall. 1975.
2. D.M. Lucantoni, M.F. Neuts, Some steady-state distribution for the $MAP/SM/1$ queue, Communications in Statistics-Stochastic Models 10 (1994) 575-598.
3. M.F. Neuts, Structured Stochastic Matrices of $M/G/1$ Type and Their Applications, Marcel Dekker, New York. 1989.

СЕТЬ С ОТРИЦАТЕЛЬНЫМИ ЗАЯВКАМИ И СИГНАЛАМИ И ОГРАНИЧЕНИЕМ НА ИХ ВРЕМЯ ПРЕБЫВАНИЯ В УЗЛАХ

О. Якубович

Гомельский государственный университет
Гомель, Беларусь
yakubovich@gsu.by

В настоящей работе исследуется модель открытой сети, в узлах которой формируются три очереди из поступающих положительных, отрицательных заявок и сигналов. Время пребывания в каждом узле поступающих требований ограничено случайной величиной, имеющей показательное распределение. Сигнал оказывает на состояние узла одно из воздействий: уменьшает длину очереди положительных заявок на единицу, увеличивает длину очереди положительных заявок на единицу или не производит никакого изменения. Определено условие эргодичности марковского процесса, описывающего функционирование сети, стационарное распределение вероятностей состояний найдено в форме произведение множителей, представляющих собой стационарные распределения изолированных узлов.

Ключевые слова: открытая сеть, положительные заявки, отрицательные заявки, сигналы, ограниченное время пребывания, условие эргодичности, стационарное распределение.

1. ВВЕДЕНИЕ

В аналитических исследованиях стационарного функционирования сетей массового обслуживания ключевым является вопрос о нахождении стационарного распределения, которое является отправной точкой большинства исследований в теории массового обслуживания. Для точного нахождения стационарного распределения многомерного процесса, описывающего поведение сети массового обслуживания, используют возможность мультипликативного представления распределения, множители которого характеризуют отдельные узлы. Требования современности диктуют необходимость развития аналитического аппарата исследования объектов, имеющих сетевую структуру, что приводит к появлению новых интересных моделей сетей массового обслуживания. Использование сложных математических моделей открывает возможности эффективного конструирования и эксплуатации исследуемых систем. В последнее время большой интерес исследователей вызывают модели с отрицательными заявками и сигналами, отличающимися от обычных положительных заявок тем, что они не нуждаются в реальном обслуживании, а

оказывают некоторое воздействие на состояние системы, например, отрицательные заявки уменьшают очередь обычных заявок непустой системы на единицу. Отрицательные заявки были впервые введены в рассмотрение Геленбе [1, 2].

В настоящей работе рассматривается модель открытой сети, в которую поступают пуассоновские потоки положительных заявок, отрицательных заявок и сигналов. Очереди в узлах формируются из положительных заявок, отрицательных заявок и сигналов. Обслуживания требуют только положительные заявки. Время пребывания в очереди узла для положительных, отрицательных заявок и сигналов ограничено случайной величиной, имеющей экспоненциальное распределение с параметром обратно пропорциональным количеству соответствующих требований, находящихся в узле. Отрицательная заявка, время пребывания в узле которой закончилось, уменьшает количество положительных заявок в узле на единицу, если такие есть в узле. Сигнал после окончания времени пребывания оказывает одно из следующих воздействий на очередь положительных заявок, находящихся в узле: уменьшает длину очереди на единицу, если очередь положительных заявок непуста, увеличивает длину очереди на единицу или не производит никакого изменения. Данное предположение может найти применение в моделировании реальных сетей связи, поскольку при передаче запроса часто устанавливается так называемый тайм-аут, истечение которого означает, что передача запроса не укладывается в планируемый интервал времени, после чего запрос удаляется из очереди. Отрицательные заявки и сигналы в рассматриваемой модели, например, могут описывать вирусы в системе, которые начинают действовать через случайное время, при этом сигнал либо „исправляется“ — становится положительной заявкой, либо „несет разрушение“ , уничтожает положмтельную заявку в системе — становится отрицательной заявкой, либо „отражается“, „уничтожается “— не оказывает воздействия на узел.

2. ИЗОЛИРОВАННЫЙ УЗЕЛ

Рассмотрим систему массового обслуживания, в которую поступает три независимых пуассоновских потока положительных заявок, отрицательных заявок и сигналов с параметрами λ^+, λ^-, λ^s соответственно. Каждая положительная заявка, поступившая в систему, увеличивает длину очереди положительных заявок в системе на единицу и требует обслуживания. Времена обслуживания положительных заявок в системе независимы, не зависят от процесса поступления и имеют показательное распределение с параметром μ. Заявки обслуживаются в порядке поступления. Каждая положительная заявка, находящаяся в системе, имеет время пребывания, ограниченное показательно распределенной случайной величиной с параметром $\nu(n) = \nu/n$, если $n \neq 0$, $\nu(n) = 0$, если $n = 0$, где n — число положительных заявок в системе, ν — некоторая положительная постоянная. Положительная заявка, которая завершила обслуживание в системе или время пребывания которой закончилось, покидает систему. Отрицательная заявка, поступившая в систему, увеличивает длину очереди отрицательных заявок на единицу. Каж-

дая отрицательная заявка, находящаяся в системе, остается в очереди случайное время, имеющее показательное распределение с параметром $\tau(m) = \tau/m$, если $m \neq 0$, $\tau(m) = 0$, если $m = 0$, где m — число отрицательных заявок в системе, τ — некоторая положительная постоянная. После окончания времени пребывания отрицательная заявка уменьшает длину очереди положительных заявок на единицу, если в системе есть положительные заявки, и не производит никаких воздействий на систему, если в системе нет положительных заявок. Сигнал, поступивший в систему, увеличивает длину очереди сигналов на единицу. Каждый сигнал, находящийся в системе, остается в очереди случайное время, имеющее показательное распределение с параметром $\omega(k) = \omega/k$, если $k \neq 0$, $\omega(k) = 0$, если $k = 0$, где k — число сигналов в системе, ω — некоторая положительная постоянная. После окончания времени пребывания в системе сигнал с вероятностью p^- уменьшает длину очереди положительных заявок на единицу, если в системе есть положительные заявки, не изменяет состояние системы, если в ней нет положительных заявок, с вероятностью p^+ увеличивает длину очереди положительных заявок на единицу, с вероятностью p^0 не производит никаких воздействий на систему. Очевидно, что $p^- + p^+ + p^0 = 1$. После окончания времени пребывания и соответствующего воздействия на систему отрицательные заявки и сигналы аннулируются.

Каждое состояние рассматриваемой системы в момент времени t будем характеризовать вектором $x(t) = (n(t), m(t), k(t))$, где $n(t)$ — число положительных заявок в системе в момент t, $m(t)$ — число отрицательных заявок в системе в момент t, $k(t)$ — число сигналов в системе в момент t . Тогда $x(t)$ — однородный марковский процесс с непрерывным временем и счетным фазовым пространством $X = \{x = (n, m, k), n, m, k = 0, 1, \ldots\}$. Пусть $\{p(x),\ x \in X\}$ — стационарное распределение вероятностей процесса $x(t)$.

Уравнения равновесия для стационарных вероятностей рассматриваемой системы имеют следующий вид:

$$p(n,m,k)\Big[\lambda^+ + \lambda^- + \lambda^s + (\mu+\nu)I_{\{n\neq 0\}} + \tau I_{\{m\neq 0\}} + \omega I_{\{k\neq 0\}}\Big] =$$
$$= p(n-1,m,k)\lambda^+ I_{\{n\neq 0\}} + p(n,m-1,k)\lambda^- I_{\{m\neq 0\}} + p(n,m,k-1)\lambda^s I_{\{k\neq 0\}} +$$
$$+ p(n+1,m,k)(\mu+\nu) + p(n+1,m+1,k)\tau + p(n,m+1,k)\tau I_{\{n=0\}} +$$
$$+ p(n-1,m,k+1)\omega p^+ I_{\{n\neq 0\}} + p(n+1,m,k+1)\omega p^- +$$
$$+ p(n,m,k+1)\omega(p^0 + p^- I_{\{n=0\}}), \quad (n,m,k) \in X.$$

Здесь $I_{\{x\}}$ —характеристическая функция, принимающая значение 1, если x истинно, и 0 — в противном случае.

Теорема 1. *Пусть выполнено условие*

$$\frac{\lambda^+ + \lambda^s p^+}{\lambda^- + \lambda^s p^- + \mu + \nu} < 1, \frac{\lambda^-}{\tau} < 1, \frac{\lambda^s}{\omega} < 1,$$

тогда марковский процесс $x(t)$ эргодичен, а финальное стационарное распределение вероятностей состояний системы имеет следующий вид

$$p(n,m,k) = \Big[\frac{\lambda^+ + \lambda^s p^+}{\lambda^- + \lambda^s p^- + \mu + \nu}\Big]^n \Big[\frac{\lambda^-}{\tau}\Big]^m \Big[\frac{\lambda^s}{\omega}\Big]^k p(0,0,0),$$

где

$$p(0,0,0) = \Big[1 - \frac{\lambda^+ + \lambda^s p^+}{\lambda^- + \lambda^s p^- + \mu + \nu}\Big]\Big[1 - \frac{\lambda^-}{\tau}\Big]\Big[1 - \frac{\lambda^s}{\omega}\Big].$$

Доказательство. Проводится стандартным образом: подстановкой стационарных вероятностей в уравнения равновесия. Условие эргодичности находится из теоремы Фостера. □

3. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ОТКРЫТОЙ СЕТИ

Рассмотрим открытую сеть, состоящую из N, узлов со структурой, описанной выше. Предположим, что в сеть поступают три независимых пуассоновских потока положительных заявок, отрицательных заявок и сигналов с параметрами λ^+, λ^-, λ^s соответственно. Каждая положительная заявка входного потока независимо от других заявок направляется в i-ый узел с вероятностью p_{0i}^+ $(i = \overline{1,N})$, $\sum\limits_{i=1}^{N} p_{0i}^+ = 1$. Каждая отрицательная заявка входного потока независимо от других направляется в i-ый узел с вероятностью p_{0i}^- $(i = \overline{1,N})$, $\sum\limits_{i=1}^{N} p_{0i}^- = 1$. Каждый сигнал входного потока независимо от других направляется в i-ый узел с вероятностью p_{0i}^s $(i = \overline{1,N})$, $\sum\limits_{i=1}^{N} p_{0i}^s = 1$. Очереди в узлах формируются из положительных, отрицательных заявок и сигналов, поступление которых увеличивает соответствующую очередь в узле на единицу. Обслуживания требуют только положительные заявки. Времена обслуживания положительных заявок независимы, не зависят от процесса поступления и в i-ом узле имеют показательное распределение с параметрами μ_i соответственно $(i = \overline{1,\ N})$. Каждая положительная заявка в i-ом узле имеет время пребывания, ограниченное экспоненциально распределенной случайной величиной с параметром $\nu_i(n_i) = \nu_i/n_i$, если $n_i \neq 0$, $\nu_i(n_i) = 0$, если $n_i = 0$, где n_i — число положительных заявок в i-ом узле, ν_i — некоторая положительная постоянная $(i = \overline{1,N})$. Положительная заявка, время пребывания которой в i-ом узле закончилось, продолжает движение по сети и ведет себя также, как положительные заявки, получившие обслуживание в i-ом узле $(i = \overline{1,N},\ k = \overline{1,\ M})$. Каждая положительная заявка, завершившая обслуживание в i-ом узле, независимо от других заявок мгновенно направляется в очередь j-го узла и становится положительной заявкой с вероятностью p_{ij}^+, или становится отрицательной заявкой с вероятностью p_{ij}^-, или становится сигналом с вероятностью p_{ij}^s, а с вероятностью p_{i0} покидает сеть. При этом $\sum\limits_{j=1}^{N}(p_{ij}^+ + p_{ij}^- + p_{ij}^s) + p_{i0} = 1,\ i = \overline{1,N}$. Предположим,

что матрица $P = (p_{ij}, i, j = \overline{0, N})$, в которой $p_{00} = 0$, $p_{i0} = p_{0i}^+$, $p_{ij} = p_{ij}^+ + p_{ij}^- + p_{ij}^s$, неприводима. Каждая отрицательная заявка в i-ом узле имеет время пребывания, ограниченное экспоненциально распределенной случайной величиной с параметром $\tau_i(m_i) = \tau_i/m_i$, если $m_i \neq 0$, $\tau_i(m_i) = 0$, если $m_i = 0$, где m_i — число положительных заявок в i-ом узле, τ_i — некоторая положительная постоянная ($i = \overline{1, N}$). Отрицательная заявка, время пребывания которой в i-ом узле закончилось, уменьшает длину очереди положительных заявок на единицу, если в узле есть положительные заявки, и не производит никаких воздействий, если в узле нет положительных заявок. Каждый сигнал, находящийся в узле, остается в очереди случайное время, имеющее показательное распределение с параметром $\omega_i(k_i) = \omega_i/k_i$, если $k_i \neq 0$, $\omega_i(k_i) = 0$, если $k_i = 0$, где k_i — число сигналов в узле, ω_i — некоторая положительная постоянная. После окончания времени пребывания в узле сигнал с вероятностью p_i^- уменьшает длину очереди положительных заявок на единицу, если в узле есть положительные заявки, не изменяет состояние узла, если в нем нет положительных заявок, с вероятностью p_i^+ увеличивает длину очереди положительных заявок на единицу, с вероятностью p_i^0 не производит никаких воздействий на узел. Очевидно, что $p_i^- + p_i^+ + p_i^0 = 1$. После окончания времени пребывания в узле и оказания соответствующего воздействия на состояние узла отрицательная заявка или сигнал аннулируются.

Обозначим через ε_i^+ среднюю интенсивность поступления положительных заявок в i-ый узел, через ε_i^- среднюю интенсивность поступления отрицательных заявок в i-ый узел, через ε_i^s среднюю интенсивность поступления сигналов в i-ый узел. Эти интенсивности удовлетворяют следующей системе нелинейных уравнений трафика:

$$\varepsilon_i^+ = \lambda^+ p_{0i}^+ + \varepsilon_i^s p_i^+ + \sum_{j=1}^{N} \frac{\varepsilon_j^+(\mu_j + \nu_j)}{\varepsilon_j^- + \varepsilon_j^s p_j^- + \mu_j + \nu_j} p_{ji}^+,$$

$$\varepsilon_i^- = \lambda^- p_{0i}^- + \sum_{j=1}^{N} \frac{\varepsilon_j^+(\mu_j + \nu_j)}{\varepsilon_j^- + \varepsilon_j^s p_j^- + \mu_j + \nu_j} p_{ji}^-,$$

$$\varepsilon_i^s = \lambda^s p_{0i}^s + \sum_{j=1}^{N} \frac{\varepsilon_j^+(\mu_j + \nu_j)}{\varepsilon_j^- + \varepsilon_j^s p_j^- + \mu_j + \nu_j} p_{ji}^s, \quad i = \overline{1, N}.$$

Доказательство существования решения нелинейных уравнений трафика проводится с помощью теоремы Брауэра о неподвижной точке [3].

Состояние рассматриваемой сети в момент времени t будем характеризовать вектором $x(t) = (x_1(t), x_2(t), \ldots, x_N(t))$, где $x_i(t) = (n_i(t), k_i(t), m_i(t))$ описывает состояние i-го узла, то есть $n_i(t)$ — число положительных заявок в i-ом узле в момент времени t, $m_i(t)$ — число отрицательных заявок в i-ом узле в момент времени t, $k_i(t)$ — число сигналов в i-ом узле в момент времени t ($i = \overline{1, N}$). Тогда $x(t)$ — однородный марковский процесс с непрерывным временем и счетным фазовым пространством $X = \{x = (x_1, x_2, \ldots, x_N), x_i = (n_i, m_i, k_i), \ n_i, m_i, k_i =$

$0, 1, 2, \ldots;\ i = \overline{1, N}\}$. Пусть $\{p(x),\, x \in X\}$ — стационарное распределение вероятностей состояний процесса $x(t)$.

Уравнения глобального равновесия для стационарных вероятностей имеют следующий вид:

$$p(x)\Big[\lambda^+ + \lambda^- + \lambda^s + \sum_{i=1}^{N}\Big((\mu_i + \nu_i)I_{\{n_i \neq 0\}} + \tau_i I_{\{m_i \neq 0\}} + \omega_i I_{\{k_i \neq 0\}}\Big)\Big] =$$

$$= \sum_{i=1}^{N}\Big[p(x + e_{i,1})(\mu_i + \nu_i)p_{i0} + p(x + e_{i,2})\tau_i I_{\{n_i = 0\}} + p(x + e_{i,3})\omega_i(p_i^- I_{\{n_i = 0\}} + p_i^0) +$$

$$+ p(x - e_{i,1})\lambda^+ p_{0i}^+ I_{\{n_i \neq 0\}} + p(x - e_{i,2})\lambda^- p_{0i}^- I_{\{m_i \neq 0\}} + p(x - e_{i,3})\lambda^s p_{0i}^s I_{\{k_i \neq 0\}} +$$

$$+ p(x + e_{i,1} + e_{i,2})\tau_i + p(x + e_{i,1} + e_{i,3})\omega_i p_i^- + p(x - e_{i,1} + e_{i,3})\omega_i p_i^+ I_{\{n_i \neq 0\}} +$$

$$+ \sum_{j=1}^{N}[p(x + e_{i,1} - e_{j,1})(\mu_i + \nu_i)p_{ij}^+ I_{\{n_j \neq 0\}} + p(x + e_{i,1} - e_{j,2})(\mu_i + \nu_i)p_{ij}^- I_{\{m_j \neq 0\}} +$$

$$+ p(x + e_{i,1} - e_{j,3})(\mu_i + \nu_i)p_{ij}^s I_{\{k_j \neq 0\}}\Big], \quad x \in X.$$

Здесь $e_{i,k}$—единичный вектор размерности $3N$ с единицей в $((i-1)N+k)$-ой позиции.

Теорема 2. *Пусть для любого $i = 1, \ldots, N$ выполнены условия*

$$\frac{\varepsilon_i^+}{\varepsilon_i^- + \varepsilon_i^s p_i^- + \mu_i + \nu_i} < 1, \quad \frac{\varepsilon_i^-}{\tau_i} < 1, \quad \frac{\varepsilon_i^s}{\omega_i} < 1,$$

тогда марковский процесс $x(t)$ эргодичен, а финальное стационарное распределение вероятностей состояний системы имеет следующий вид

$$p(x) = p_1(x_1)p_2(x_2)\ldots p_N(x_N), \quad x \in X,$$

где

$$p_i(x_i) = \Big[\frac{\varepsilon_i^+}{\varepsilon_i^- + \varepsilon_i^s p_i^- + \mu_i + \nu_i}\Big]^{n_i}\Big[\frac{\varepsilon_i^-}{\tau_i}\Big]^{m_i}\Big[\frac{\varepsilon_i^s}{\omega_i}\Big]^{k_i} p_i(0,0,0),$$

$$p_i(0,0,0) = \Big[1 - \frac{\varepsilon_i^+}{\varepsilon_i^- + \varepsilon_i^s p_i^- + \mu_i + \nu_i}\Big]\Big[1 - \frac{\varepsilon_i^-}{\tau_i}\Big]\Big[1 - \frac{\varepsilon_i^s}{\omega_i}\Big],$$

$(\varepsilon_i^+, \varepsilon_i^-, \varepsilon_i^s,\ i = \overline{1, N})$ *находятся как решение нелинейных уравнений трафика.*

Доказательство теоремы проводится стандартным образом, подстановкой стационарного распределения в уравнения равновесия.

ЛИТЕРАТУРА

1. *Gelenbe E.* Product-form networks with negative and positive customers // J. Appl. Prob. 1991. V. 28. P. 656–663.
2. *Gelenbe E., Schassberger R.* Stability of product-form G-networks // Probab. Eng. and Inf. Sci. 1992. V. 6. P. 271–276
3. *Данфорд Н., Шварц Дж. Т.* Линейные операторы. Общая теория. М.:ИЛ, 1962.

ON STABILITY FOR NONSTATIONARY $M/M/N/N+R$ QUEUE

A. Zeifman[1], A. Korotysheva[2], T. Panfilova[2], Ya. Satin[2], G. Shilova[2]

[1] *Vologda State Pedagogical University, Institute of Informatics Problems RAS, and ISEDT RAS,*
[2] *Vologda State Pedagogical University*
[1],[2] *Vologda, Russia*
a_zeifman@mail.ru

We consider nonstationary $M/M/N/N+R$ queueing model and obtain the first bounds of stability.

Keywords: Nonstationary queues; $M_t/M_t/N/N+R$ queueing model; bounds; stability.

1. INTRODUCTION

In this note we consider the estimates of stability for the simplest generalization of nonstationary Erlang queueing model. There is a number of investigations of nonstationary continuous-time Markov chains, see for instance first results in [5], and more detail studies for birth and death processes (BDPs) in [1, 6]. Now we consider nonstationary $M_t/M_t/N/N+R$ queue and obtain some simple stability bounds.

Let $X = X(t)$, $t \geq 0$ be queue-length process for $M_t/M_t/N/N+R$ queue. This is a BDP on state space $E_{N+R} = \{0, 1 \ldots, N+R\}$ and birth and death rates $\lambda_n(t) = \lambda(t)$, $\mu_n(t) = \min(n, N)\,\mu(t)$ respectively. We suppose that arrival and service intensities $\lambda(t)$ and $\mu(t)$ are locally integrable on $[0, \infty)$. Let $p_i(t) = Pr\{X(t) = i\}$ be state probabilities of $X(t)$, and $\mathbf{p}(t) = (p_0(t), \ldots, p_{N+R}(t))^T$ be the respective column vector.

Then we can write the forward Kolmogorov system

$$\left\{\begin{array}{c} \frac{dp_0}{dt} = -\lambda(t)p_0 + \mu(t)p_1, \\ \frac{dp_k}{dt} = \lambda(t)p_{k-1} - (\lambda(t) + k\mu(t))\,p_k + (k+1)\mu(t)p_{k+1}, 1 \leq k \leq N-1, \\ \frac{dp_k}{dt} = \lambda(t)p_{k-1} - (\lambda(t) + N\mu(t))\,p_k + N\mu(t)p_{k+1}, N \leq k < N+R, \\ \frac{dp_N}{dt} = \lambda(t)p_{N-1} - N\mu(t)p_N \end{array}\right. \tag{1}$$

in the following form:

$$\frac{d\mathbf{p}}{dt} = A\,(t)\,\mathbf{p}, \quad t \geq 0, \tag{2}$$

where $A(t) = \{a_{ij}(t),\ t \geq 0\}$ is the transposed intensity matrix of the process, and

$$a_{ij}(t) = \begin{cases} \lambda(t), & \text{if} \quad j = i-1, \\ \min(i+1, N)\,\mu(t), & \text{if} \quad j = i+1, \\ -(\lambda(t) + \min(i, N)\,\mu(t)), & \text{if} \quad j = i, \\ 0, & \text{overwise.} \end{cases} \tag{3}$$

We denote throughout the paper by $\| \bullet \|$ the l_1-norm, i.e. $\|\mathbf{x}\| = \sum |x_i|$, for $\mathbf{x} = (x_0, ..., x_{N+R})^T$ and $\|B\| = \max_j \sum_i |b_{ij}|$ for $B = (b_{ij})_{i,j=0}^{N+R}$.

Let $\Omega = \{\mathbf{x} :\ \mathbf{x} \geq 0,\ \|\mathbf{x}\| = 1\}$ be a set of all stochastic vectors.

Let $E_k(t) = E\{X(t)\,|X(0) = k\}$ be the mean of the process at the moment t under initial condition $X(0) = k$, and $E_{\mathbf{p}}(t)$ be the mathematical expectation (the mean) at the moment t under initial probability distribution $\mathbf{p}(0) = \mathbf{p}$.

Consider also a "perturbed" queue-length process $\bar{X} = \bar{X}(t),\ t \geq 0$ with general structure of intensity matrix $\bar{A}(t)$. In general, $\bar{X}(t)$ is not BDP. Put $\hat{A}(t) = \bar{A}(t) - A(t)$. We assume that the perturbations are uniformly small, i.e. $\|\hat{A}(t)\| \leq \varepsilon$ for almost all $t \geq 0$.

2. STABILITY BOUNDS

Let $d_1, \ldots, d_{N+R}$ be positive numbers. Consider the following expression:

$$\alpha_i(t) = \lambda(t) + \min(i, N)\,\mu(t) - \frac{d_{i+1}}{d_i}\lambda(t) - \frac{d_{i-1}}{d_i}\min(i-1, N)\,\mu(t),$$
$$i = 1, 2, \ldots, N+R, \tag{4}$$

where $d_0 = d_{N+R+1} = 0$. Put $G = \sum_{i=1}^{N+R} d_i$ and $d = \min_{1 \leq i \leq N+R} d_i$.

Theorem 1. *Let there exist a positive sequence $\{d_i\}$ and a positive number θ such that*

$$\alpha_i(t) \geq \theta, \quad i = 1, 2, \ldots, N+R,\ t \geq 0. \tag{5}$$

Then the following stability bounds hold:

$$\limsup_{t \to \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{\varepsilon\left(1 + \log\frac{4G}{d}\right)}{\theta}, \tag{6}$$

and

$$\limsup_{t \to \infty} \left|E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}}(t)\right| \leq \frac{(N+R)\,\varepsilon\left(1 + \log\frac{4G}{d}\right)}{\theta}, \tag{7}$$

for arbitrary initial probability distributions $\mathbf{p}(0)$ and $\bar{\mathbf{p}}(0)$ for $X(t)$ and $\bar{X}(t)$ respectively.

Proof. Firstly we find the basic estimate of the rate of convergence. The property $\sum_{i=0}^{N+R} p_i(t) = 1$ for any $t \geq s$ allows to put $p_0(t) = 1 - \sum_{i \geq 1} p_i(t)$, then we obtain the following system from (2)

$$\frac{d\mathbf{z}(t)}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t), \tag{8}$$

where $\mathbf{z}(t) = (p_1(t), \ldots, p_{N+R}(t))^T$, $\mathbf{f}(t) = (\lambda(t), 0, \ldots, 0)^T$, $B(t) = (b_{ij}(t))_{i,j=1}^{N+R}$ and respective $b_{ij}(t)$, see details in [7, 8]. Consider now the triangular matrix

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots & d_1 \\ 0 & d_2 & d_2 & \cdots & d_2 \\ 0 & 0 & d_3 & \cdots & d_3 \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & 0 & 0 & d_{N+R} \end{pmatrix}, \tag{9}$$

and the respective norms $\|\mathbf{x}\|_{1D} = \|D\mathbf{x}\|$, and $\|B\|_{1D} = \|DBD^{-1}\|$.

We have now the following bound of the logarithmic norm $\gamma(B(t))$ in $1D-$norm (see for instance [2, 3, 7, 9]):

$$\begin{aligned} \gamma(B)_{1D} = \max_i \Big(\frac{d_{i+1}}{d_i}\lambda(t) + \frac{d_{i-1}}{d_i} \min(i-1, N)\, \mu(t) - \\ (\lambda(t) + \min(i, N)\, \mu(t))) = \max(-\alpha_i(t)) \le -\theta, \end{aligned} \tag{10}$$

in accordance with (5). Therefore the following inequality holds:

$$\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\|_{1D} \le e^{-\theta(t-s)} \|\mathbf{z}^*(s) - \mathbf{z}^{**}(s)\|_{1D}, \tag{11}$$

for any initial conditions $\mathbf{z}^*(s)$, $\mathbf{z}^{**}(s)$ and any $s, t,\ 0 \le s \le t$. Then we obtain

$$\begin{aligned} \|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \le 2\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\| = \\ 2\|D^{-1}D\left(\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\right)\| \le \\ \frac{4}{d}\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\|_{1D} \le \\ \frac{4}{d}e^{-\theta(t-s)}\|\mathbf{z}^*(s) - \mathbf{z}^{**}(s)\|_{1D} \le \\ \frac{4G}{d}e^{-\theta(t-s)}\|\mathbf{z}^*(s) - \mathbf{z}^{**}(s)\| \le \\ \frac{4G}{d}e^{-\theta(t-s)}\|\mathbf{p}^*(s) - \mathbf{p}^{**}(s)\| \le \frac{8G}{d}e^{-\theta(t-s)}, \end{aligned} \tag{12}$$

for any initial conditions $\mathbf{p}^*(s)$, $\mathbf{p}^{**}(s)$ and any $s, t,\ 0 \le s \le t$.

Consider the forward Kolmogorov system for perturbed process:

$$\frac{d\bar{\mathbf{p}}}{dt} = \bar{A}(t)\bar{\mathbf{p}}(t). \tag{13}$$

We can apply the approach of paper [4]. Put

$$\beta(t,s) = \sup_{\|\mathbf{v}\|=1, \sum v_i = 0} \|U(t,s)\mathbf{v}\| = \frac{1}{2} \max_{i,j} \sum_k |p_{ik}(t,s) - p_{jk}(t,s)|, \tag{14}$$

where $U(t,s)$ is Cauchy matrix of (2), and $p_{ik}(t,s) = Pr\{X(t) = k | X(s) = i\}$. Mitrophanov in [4] proved the bound of stability, that in the nonstationary case is the following one:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \le \beta(t,s)\|\mathbf{p}(s) - \bar{\mathbf{p}}(s)\| + \int_s^t \|\hat{A}(u)\|\beta(u,s)\,du. \tag{15}$$

Moreover, the following estimates hold:

$$\beta(t,s) \le 1, \quad \beta(t,s) \le \frac{ce^{-b(t-s)}}{2}, \ 0 \le s \le t, \tag{16}$$

where under our assumptions $c = \frac{8G}{d}$, $b = \theta$. Finally the following stability bound holds:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \le \begin{cases} \|\mathbf{p}(s) - \bar{\mathbf{p}}(s)\| + (t-s)\varepsilon, \\ 0 < t - s < b^{-1}\log\frac{c}{2}, \\ b^{-1}(\log\frac{c}{2} + 1 - ce^{-b(t-s)})\varepsilon + \\ \frac{c}{2}e^{-b(t-s)}\|\mathbf{p}(s) - \bar{\mathbf{p}}(s)\|, \quad t - s \ge b^{-1}\log\frac{c}{2}, \end{cases} \tag{17}$$

for any initial conditions $\mathbf{p}(s)$, $\bar{\mathbf{p}}(s)$. Let $t - s \to \infty$. Then (17) implies our claim. □

Consider here the case of sufficiently large service rate, namely let there exist $d > 1$ such that the following assumption holds:

$$N\mu(t) - d\lambda(t) \ge \theta^* > 0, \tag{18}$$

for any $t \ge 0$.

Put $d_1 = 1$, $\frac{d_{k+1}}{d_k} = \delta_k = 1$, $k \le N-2$, and $\frac{d_{k+1}}{d_k} = \delta_k = d$, $k \ge N-1$.

Then

$$\alpha_k(t) = \begin{cases} \mu(t), & k < N-1; \\ \mu(t) - (d-1)\lambda(t), & k = N-1; \\ \left(1 - \frac{1}{d}\right)(N\mu(t) - d\lambda(t)), & N \le k \le N+R-2; \\ N\mu(t)\left(1 - \frac{1}{d}\right) - \lambda(t), & k = N+R-1. \end{cases} \tag{19}$$

Let $d \le \frac{N}{N-1}$, then we obtain

$$\theta = \inf_k \alpha_k(t) = \left(1 - \frac{1}{d}\right)(N\mu(t) - d\lambda(t)) \ge \left(1 - \frac{1}{d}\right)\theta^*. \tag{20}$$

Hence we obtain the following statement.

Theorem 2. *Under assumption (18) stability bounds (6) and (7) hold for* $\theta = \left(1 - \frac{1}{d}\right)\theta^*$, $d = 1$, *and* $G = N - 1 + \sum_{k=1}^{R+1} d^k$.

Let now our original process have 1-periodic intensities.

Then the following claim holds.

Theorem 3. *Let* $\lambda(t)$ *and* $\mu(t)$ *be 1-periodic. Let there exist a positive sequence* $\{d_i\}$ *and a positive number* φ^* *such that*

$$\alpha_i(t) \geq \varphi(t), \quad i = 1, 2, \ldots, N + R, 0 \leq t \leq 1, \tag{21}$$

where

$$\int_0^1 \varphi(t)\, dt \geq \varphi^*. \tag{22}$$

Let

$$K = \sup_{|t-s| \leq 1} \int_s^t \varphi(\tau)\, d\tau < \infty. \tag{23}$$

Then we have the following stability bounds:

$$\limsup_{t \to \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{\varepsilon \left(1 + \log \frac{4Ge^K}{d}\right)}{\varphi^*}, \tag{24}$$

and

$$\limsup_{t \to \infty} \left|E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}}(t)\right| \leq \frac{(N + R)\, \varepsilon \left(1 + \log \frac{4Ge^K}{d}\right)}{\varphi^*}, \tag{25}$$

for arbitrary initial probability distributions $\mathbf{p}(0)$ *and* $\bar{\mathbf{p}}(0)$ *for* $X(t)$ *and* $\bar{X}(t)$ *respectively.*

Proof. The statement follows from inequality $e^{-\int_s^t \varphi(u)\, du} \leq e^K e^{-\varphi^*(t-s)}$. □

In the case of sufficiently large service rate we obtain the following claim.

Theorem 4. *Let arrival and service rates be 1-periodic, and let*

$$\int_0^1 (N\mu(\tau) - d\lambda(\tau))\, d\tau = \psi > 0, \tag{26}$$

instead of (18). Then stability bounds (24) and (25) hold for $\varphi^* = \left(1 - \frac{1}{d}\right)\psi$, $d = 1$, *and* $G = N - 1 + \sum_{k=1}^{R+1} d^k$.

REFERENCES

1. *Andreev D. B. at al.* Ergodicity and stability of nonstationary queueing systems // Theory of Probability and Math. Statist. 2004. V. 68. P. 1–10.
2. *Van Doorn E. A., Zeifman A. I., Panfilova T. L.* Bounds and asymptotics for the rate of convergence of birth-death processes // Theory of Probability and Its Applications. 2010. V. 54. P. 97–113.
3. *Granovsky B., Zeifman A.* Nonstationary queues: Estimation of the rate of convergence // Queueing Syst. 2004. V. 46. P. 363–388.
4. *Mitrophanov A. Yu.* Stability and exponential convergence of continuous-time Markov chains // J. Appl. Prob. 2003. V. 40. P. 970–979.
5. *Zeifman A. I.* Stability for contionuous-time nonhomogeneous Markov chains // Lect. Notes Math. 1985. V. 1155. P. 401–414.
6. *Zeifman A.* Stability of birth and death processes // Journal of Mathematical Sciences. 1998. V. 91. P. 3023–3031.
7. *Zeifman A., Leorato S., Orsingher E., Satin Ya., Shilova G.* Some universal limits for nonhomogeneous birth and death processes // Queueing Syst. 2006. V. 52. P.139–151.
8. *Zeifman A. I., Bening V. E., Sokolov I. A.* Continuous-time Markov chains and models (in Russian) // Moscow, Elex-KM, 2008.
9. Zeifman A. I. On the nonstationary Erlang loss model // Automation and Remote Control. 2009. V. 70. P. 2003–2012.

COST OF REACHING WITH PROHIBITION FOR AN EXPONENTIAL TIME-SHARING QUEUEING PROCESS

A. Zorine

N.I. Lobachevsky State University
of Nizhni Novgorod
National Research University
Nizhni Novgorod, Russia
zoav1@uic.nnov.ru

A process of servicing of conflict flows with a time-sharing algorithm with readjustments is considered. Input flows are formed in a random environment synchronized with the servicing device. Service and readjustment durations are exponentially distributed. A mathematical model for the servicing process is constructed in form of a denumerable continuous-time Markov chain. To estimate the loading degree of the time-sharing queueing system with readjustments an expected cost of reaching with prohibition is proposed. It generalizes Chung functionals for homogeneous denumerable discrete-time Markov chains. A control is searched for which minimizes the functional in a class of simple priority controls with threshold.

Keywords: time-sharing service with readjustments, random environment, loading of a queueing system, Chung functionals, denumerable continuous-time Markov chains.

1. INTRODUCTION

It is important for practice to be able to estimate the loading of a queueing system. As a rule a loading is understood as a fraction of time which the system spends on servicing of customers. However this definition is not that unambiguous for a number of real queueing systems, because it does not take into account a possibility of fixed cycles in servicing of queues (e.g. in servicing of traffic flows) or presence of readjustment-and-control periods in servicing of conflict flows (queueing systems controlling traffic flows, time-sharing information processing systems, etc.). In particular, no explicit expression is known for the loading of a time-sharing queueing system with readjustments in random environment [1].

In a series of papers a reaching with prohibitions functional is used for estimation of the system's loading. Such a functional is called a *Chung functional* ([2]). In the present paper we extend the notion of a Chung functional to a class of denumerable continuous-time Markov chains.

2. PROBLEM STATEMENT

There are $m < \infty$ input flows Π_1, Π_2, $\ldots$, Π_m of customers entering into the queueing system. The flows are formed in a random external environment with two states, $e^{(1)}$ and $e^{(2)}$. During the environmental state $e^{(k)}$, $k = 1, 2$, the flow Π_j, $j = 1, 2, \ldots, m$, is poissonian with intensity $\lambda_j^{(k)}$. Customers from the flow Π_j are put into an infinite buffer (a queue) O_j. Customers are selected for service in order of arrival. A servicing device has $n = 2m + 1$ states, $\Gamma^{(0)}$, $\Gamma^{(1)}$, $\ldots$, $\Gamma^{(2m)}$. In a state $\Gamma^{(r)}$, $1 \leqslant r \leqslant m$, a customer from the queue O_r is serviced. After the state $\Gamma^{(r)}$ the servicing device switches into a state $\Gamma^{(r')}$ with $r' = r + m$, $m + 1 \leqslant r' \leqslant 2m$. During the state $\Gamma^{(r')}$ a readjustment-and-control act is done and no customers are serviced. If at an readjustment-and-control act termination the queues are empty the servicing device switches into the state $\Gamma^{(0)}$ of waiting for arrivals. Upon an arrival of the first customer its service begins immediately and the server switches to a state $\Gamma^{(j)}$ if the customer is from the flow $\Pi^{(j)}$. If the queues' lengths are described with a nonzero vector $x = (x_1, x_2, \ldots, x_m)$ after a readjustment-and-control act termination then a customer from the queue O_j, $j = h(x)$, is selected for service where $h(\cdot)$ is a given mapping of the non-negative integer lattice $X = \{0, 1, \ldots\}^m$ onto the set $\{0, 1, \ldots, m\}$. The mapping must satisfy to the following constraints: $h(x) = j$ implies $x_j > 0$ and only the zero vector $\bar{0} = (0, 0, \ldots, 0) \in X$ is mapped into 0. Service durations and readjustment durations are independent random variables with exponential distributions. The mean service time at the state $\Gamma^{(r)}$ is β_r^{-1}, and the mean readjustment-and-control time at the state $\Gamma^{(m+r)}$ equals $\bar{\beta}_r^{-1}$. The random external environment is synchronized with the servicing device. Change of the environmental state can occur only at instant of service or readjustment-and-control acts terminations. A probability of changing the state from $e^{(k)}$ to $e^{(l)}$ equals $a_{k,l}$, k, $l = 1, 2$. Thus the process of environmental state changing is not markovian. After a service a customer from the queue O_j can get into the queue O_r with probability $p_{j,r}$ for a repeated service, and can quit the queueing system with probability $p_{j,m+1} = 1 - \sum_{r=1}^{m} p_{j,r}$. So, besides the flows of primary customers there are flows of secondary customers entering the queueing system. The total input flows have a complicated probabilistic structure. A sojourn cost for one customer in O_j per time unit is given and it equals c_j. Suppose a decomposition of the set X of vectors of queues' lengths info non-empty disjoint subsets X_0, X_+, X_-, $X = X_0 \cup X_+ \cup X_-$ is defined. The set X_0 is interpreted as a set of admissible queues' lengths, the set X_+ stands for a set of desired queues' lengths, finally, the set X_- consists of prohibited queues' lengths. In essence the aim of the present paper is to give a mathematical formulation of an optimization problem and to study numerically an algorithms $h(\cdot)$ which minimizes the conditional expected cost of reaching the desirable set X_+ from the admissible set X_0 under hypothesis of not visiting the prohibited set X_-.

The queueing system described can be obtained as a special case of a time-sharing queueiny system in [1] with exponentially distributed service and readjustment durations. It was shown in the cited paper that a priority algorithm minimizes the expected sojourn cost of all customers int the system during an act in the stationary mode.

3. MATHEMATICAL MODEL AND OPTIMIZATION PROBLEM

In what follows all objects are defined or constructed on a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ where Ω is a set of descriptions ω of elementary outcomes, $\mathfrak{F}$ is a σ-algebra of events $A \subset \subset \Omega$, $\mathbf{P}$ is a probability. Introduce the following random variables and random elements. Let $\chi(t) \in \{e^{(1)}, e^{(2)}\}$ be the state of the environment at the instant $t \geqslant 0$, $\Gamma(t) \in \Gamma = = \{\Gamma^{(0)}, \Gamma^{(1)}, \ldots, \Gamma^{(2m)}\}$ the state of the servicing device at the instant t, $\varkappa_j(t)$ the number of customers in the queue O_j at the instant t, $\varkappa(t) = (\varkappa_1(t), \varkappa_2(t), \ldots, \varkappa_m(t))$.

Since the evolution of the process

$$\{(\Gamma(t), \varkappa(t), \chi(t)); t \geqslant 0\} \tag{1}$$

after the instant t is determined only by the instant of termination of an act taking place at t, by arrivals of customers after the instant t, and by the states of the random external environment after the instant t, the random process (1) is markovian with denumerable number of states and a states space

$$\begin{aligned} S = \{(\Gamma^{(0)}, \bar{0}, e^{(k)})\colon k = 1, 2\} \cup \\ \cup \{(\Gamma^{(r)}, x, e^{(k)})\colon r = 1, 2, \ldots, m;\ x \in X,\ x_r > 0;\ k = 1, 2\} \cup \\ \cup \{(\Gamma^{(r)}, x, e^{(k)})\colon r = m+1, m+2, \ldots, 2m;\ x \in X;\ k = 1, 2\}. \end{aligned}$$

The Markov process (1) is homogeneous. Denote by $q(r, x, k; r', w, l)$ the intensity of transition from $(\Gamma^{(r)}, x, k) \in S$ into $(\Gamma^{(r')}, w, l) \in S$. Put $q(r, x, k) = -q(r, x, k; r, x, k)$, let $\delta_{s,s'}$ be the Kronecker delta. For $s = 1, 2, \ldots, m+1$ introduce non-random vectors $y^{(s)} = (\delta_{1,s}, \delta_{2,s}, \ldots, \delta_{m,s}) \in X$, where $y^{(m+1)} = \bar{0}$.

Theorem 1. *The transition intensities of the Markov process (1) are*

$$q(0, \bar{0}, k) = \lambda_1^{(k)} + \lambda_2^{(k)} + \ldots + \lambda_m^{(k)},$$

for $r = 1, 2, \ldots, m$

$$q(r, x, k) = \beta_r^{-1} + \lambda_1^{(k)} + \lambda_2^{(k)} + \ldots + \lambda_m^{(k)},$$

for $r = m+1, m+2, \ldots, 2m$

$$q(r, x, k) = \bar{\beta}_r^{-1} + \lambda_1^{(k)} + \lambda_2^{(k)} + \ldots + \lambda_m^{(k)},$$

for $r = 1, 2, \ldots, m$

$$q(0, \bar{0}, k; r, y^{(r)}, k) = \lambda_r^{(k)},$$

for $r = 1, 2, \ldots, m$, $x \in X$, $x_r > 0$, $j = 1, 2, \ldots, m$

$$q(r, x, k; r, x + y^{(j)}, k) = \lambda_j^{(k)},$$

$$q(r, x, k; m + r, x - y^{(r)} + y^{(s)}, l) = \beta_r^{-1} p_{r,s} a_{k,l},$$

for $r = m+1,\ m+2,\ \ldots,\ 2m,\ j = 1,\ 2,\ \ldots,\ m,\ r' = h(x)$

$$q(r,x,k;r,x+y^{(j)},k) = \lambda_j^{(k)},$$
$$q(r,x,k;r',x,l) = \bar{\beta}_r^{-1} a_{k,l}.$$

Remaining intensities are equal to zero.

It follows from the form of the transition probabilities that the Markov process (1) has only stable non-absorbing states and a conservative infinitesimal matrix.

Let $\tau_0 = 0$ and $\tau_1,\ \tau_2,\ \ldots$ be the jump instants of the process (1). Consider the embedded Markov chain

$$\{(\Gamma_i, \varkappa_i, \chi_i); i = 0, 1, \ldots\} \tag{2}$$

with $\Gamma_i = \Gamma(\tau_i + 0)$, $\varkappa_i = \varkappa(\tau_i + 0)$, $\chi_i = \chi(\tau_i + 0)$. It is well known that for the Markov chain (2) the probability of transition from a state $(r,x,k) \in S$ into a state $(r',w,l) \in S$ equals $q(r,x,k;r',w,l)(q(r,x,k))^{-1}$ for $(r,x,k) \neq (r',w,l)$ and equals 0 for $(r,x,k) = (r',w,l)$, while the time interval $\tau_{i+1} - \tau_i$ given $(\Gamma_i, \varkappa_i, \chi_i) = (\Gamma^{(r)}, x, e^{(k)})$ has an exponential distribution with parameter $q(r,x,k)$. Denote

$$S_0 = \{(\Gamma^{(r)}, x, e^{(k)}) \colon (\Gamma^{(r)}, x, e^{(k)}) \in S, x \in X_0\},$$
$$S_+ = \{(\Gamma^{(r)}, x, e^{(k)}) \colon (\Gamma^{(r)}, x, e^{(k)}) \in S, x \in X_+\},$$
$$S_- = \{(\Gamma^{(r)}, x, e^{(k)}) \colon (\Gamma^{(r)}, x, e^{(k)}) \in S, x \in X_-\}.$$

For a state $(\Gamma^{(r)}, x, e^{(k)}) \in S_0$ define on the set $\Omega(r,x,k) = \{\omega \colon \Gamma_0 = \Gamma^{(r)}, \varkappa_0 = x, \chi_0 = e^{(k)}\}$ an *instant of reaching with prohibition* $\theta(r,x,k) = \min\{i \colon \varkappa_i \in S_+, \varkappa_{\bar{\imath}} \notin S_-, 0 \leqslant \bar{\imath} \leqslant i\}$. For $i = 1, 2, \ldots$ and a sequence of states $(\Gamma^{(r_{\bar{\imath}})}, x^{(\bar{\imath})}, e^{(k_{\bar{\imath}})}) \in S_0$, $0 \leqslant \bar{\imath} < i$ and $(\Gamma^{(r_i)}, x^{(i)}, e^{(k_i)}) \in S_+$ introduce a random variable $\zeta(r_0, x^{(0)}, k_0)$ which takes on value

$$\sum_{\bar{\imath}=0}^{i-1} (q(r_{\bar{\imath}}, x^{(\bar{\imath})}, k_{\bar{\imath}}))^{-1} (c_1 x_1^{(\bar{\imath})} + c_2 x_2^{(\bar{\imath})} + \ldots + c_m x_m^{(\bar{\imath})})$$

on the set $\{\omega \colon \Gamma_{\bar{\imath}} = \Gamma^{(r_{\bar{\imath}})}, \varkappa_{\bar{\imath}} = x^{(\bar{\imath})}, \chi_{\bar{\imath}} = e^{(k_{\bar{\imath}})}, 0 \leqslant \bar{\imath} \leqslant i\}$. A variable $\zeta(r,x,k)$ defines the *cost of reaching with prohibition* of the set S_+ from a state $(\Gamma^{(r)}, x, e^{(k)}) \in S_0$ without visiting the prohibited set S_-. It is interesting to investigate the conditional expected cost of reaching with prohibition

$$\mathbf{E}(\zeta(r,x,k) | \Omega(r,x,k) \cap \{\omega \colon \theta(r,x,k) < \infty\}).$$

Recall that the solution to a system of linear algebraic equations obtained by successive iterations with zero initial condition is called its *main solution* [3].

Theorem 2. *Probabilities* $f(r,x,k) = \mathbf{P}(\{\omega \colon \theta(r,x,k) < \infty\} | \Omega(r,x,k))$ *of reaching with prohibition,* $(\Gamma^{(r)}, x, e^{(k)}) \in S_0$, *are the main solution to the system of linear algebraic equations*

$$q(r,x,k) f(r,x,k) = \sum_{(\Gamma^{(r')}, w, e^{(l)}) \in S_+} q(r,x,k;r',w,l) + \\ + \sum_{(\Gamma^{(r')}, w, e^{(l)}) \in S_0} q(r,x,k;r',w,l) f(r',w',l). \tag{3}$$

The conditional expected costs of reaching with prohibition are

$$\mathbf{E}(\zeta(r,x,k)|\Omega(r,x,k)\cap\{\omega\colon\theta(r,x,k)<\infty\}) = G(r,x,k)(f(r,x,k))^{-1},$$

where the quantities $G(r,x,k)$, $(\Gamma^{(r)},x,e^{(k)})\in S_0$, *are the main solution to the system of linear algebraic equations*

$$q(r,x,k)G(r,x,k) = f(r,x,k)\sum_{j=1}^{m}c_jx_j + \sum_{(\Gamma^{(r')},w,e^{(l)})\in S_0} q(r,x,k;r',w,l)G(r',w',l). \quad (4)$$

Theorem 2 generalizes theorems in [2], which characterized the conditional expected number of steps for reaching S_+ from the states in S_0 with prohibition by a denumerable Markov chain. Equations (3), (4) allow to calculate the conditional expected costs of reaching with prohibition given the switching function $h(\cdot)$, for any initial state in S_0. Regarding the conditional expected costs as objective functionals we have a multiobjective optimization problem. To formulate an optimization problem with one objective we have to introduce the convolution of the original objectives. Define

$$J(h) = \frac{1}{|S_0|}\sum_{(\Gamma^{(r)},x,e^{(k)})\in S_0} G(r,x,k)(f(r,x,k))^{-1}. \quad (5)$$

Given a decomposition $S = S_-\cup S_0\cup S_+$ the optimization problem is reduced to solving

$$J(h_0) = \min_{h(\cdot)} J(h).$$

If the decomposition of the states space S was generated, e. g., by sets $X_+ = \{x\colon x_j \leqslant \leqslant N_j, j=1,2,\ldots,m\}$, $X_- = \{x\colon (\exists j)x_j > N_j'\}$, or $X_+ = \{x\colon x_1+x_2+\ldots+x_m \leqslant N_1\}$, $X_- = \{x\colon x_1+x_2+\ldots+x_m > N\}$, for certain positive integers N_j, N_j', $j=1,2,\ldots,m$, then the corresponding optimization problem can be understood as a problem of minimization of the conditional expected cost of unloading the queueing system since the set X_+ in essence consists of vectors of small queues' lengths.

4. SOME NUMERICAL RESULTS

To carry out numerical experiments we have to bound ourselves with a class of switching functions of a simple form since solving the optimization problem by exhaustive search is practically unachievable. Assume $m=2$ input flows. Introduce the following switching functions. Let $x=(x_1,x_2)$ and $a=0,1,\ldots$. Put

$$h_{max}(x) = \begin{cases} 0, & x=\bar{0},\\ 1, & x_1\geqslant x_2,\\ 2, & x_1<x_2;\end{cases}$$

$$h_{\mathrm{pri},1}(x) = \begin{cases} 0, & x=\bar{0},\\ 1, & x_1>0,\\ 2, & x_1=0, x_2>0;\end{cases}$$

$$h_{\text{pri},2}(x) = \begin{cases} 0, & x = \bar{0}, \\ 1, & x_2 > 0, \\ 2, & x_2 = 0, x_1 > 0; \end{cases}$$

$$h_{\text{thr},1}(x) = \begin{cases} 0, & x = \bar{0}, \\ 1, & x_2 = 0 \text{ or } x_1 \geqslant a, \\ 2, & x_2 > 0 \text{ and } x_1 < a; \end{cases}$$

$$h_{\text{thr},2}(x) = \begin{cases} 0, & x = \bar{0}, \\ 1, & x_1 > 0 \text{ and } x_2 < a, \\ 2, & x_1 = 0 \text{ or } x_2 \geqslant a. \end{cases}$$

The switching function $h_{\max}(\cdot)$ corresponds to servicing of the longest queue, $h_{\text{pri},1}(\cdot)$ and $h_{\text{pri},2}(\cdot)$ correspond to service with relative priorities, $h_{\text{thr},1}(x)$ and $h_{\text{thr},2}(x)$ correspond to a priority service with positive threshold value a for the high priority queue. Assume the following parameters are given: $a_{1,1} = 0.5$, $a_{2,2} = 0.8$, $\lambda_1^{(1)} = 0.1$, $\lambda_1^{(2)} = 0.2$, $\lambda_2^{(1)} = 0.15$, $\lambda_2^{(2)} = 0.05$, $\beta_1 = 0.375$, $\beta_2 = 0.5$, $\bar{\beta}_1 = 0.125$, $\bar{\beta}_2 = 0.25$, $p_{1,1} = 0$, $p_{1,2} = 0.05$, $p_{2,1} = 0.02$, $p_{2,2} = 0.01$, $c_1 = c_2 = 1$, $N_1 = 5$, $N_1' = 15$, $N_2 = 4$, $N_2' = 10$. The choice of costs c_1, c_2 gives us the total sojourn time of all customers in the system until unloading. Then $J(h_{\max}) = 205.779$, $J(h_{\text{pri},1}) = 237.444$, $J(h_{\text{pri},2}) = 239.188$, and the smallest value 193.852 is achieved by $J(\cdot)$ for the switching function $h_{\text{thr},1}$ with $a = 6$. Although we haven't found the optimal switching function among all possible switching functions, we still can conclude that a priority algorithm (which was known to be optimal to minimize total sojourn cost of all customers present in the queueing system per tact, cf. [1]) is not the best way to unload a time-sharing queueing system with readjustments at least with the decomposition of X considered here.

REFERENCES

1. Fedotkin M. A., Zorin A. V. *Optimization of Control of Doubly Stochastic Nonordinary Flows in Time-Sharing Systems* // Automation and remote control, 2005. V. 66, №7. P. 1115–1124.
2. Fedotkin M. A. *Algebraic properties of distributions for Chung functionals of homogeneous Markov chains with a denumerable set of states* // Sov. Math., Dokl., 1976. V. 17, P. 43–46.
3. *Kantorovich L. V., Krylov V. I. Approximate methods in higher analysis*. L.: Fizmatlit, 1962.

АВТОРСКИЙ УКАЗАТЕЛЬ
THE AUTHOR'S INDEX

Aissani A. 191
Aïssani D. 38
Ананина И. 8
Andronov A. 14
Arrar K.N. 49
Avrachenkov K. 22
Babitski A. 29
Baillon J.B. 49
Bojarovich J. 34
Boualem M. 38
Chechelnitsky A. 133
Dimitrov M. 44
Djellab N. 38, 49
Dragieva V. 52
Dudin A. 58, 258
Дудин С. 65
Ерёмина А. 71
Федоткин М. 76, 82
Goricheva R. 22
Горцев А. 88
Ivanova M. 44
Калягин А. 96
Ким Ч.С. 102
Кищенко М. 121
Kiturko O. 108
Klimenok V. 58, 102, 114, 258
Колузаева Е. 121, 196
Korotysheva A. 271
Кудрявцев Е. 76
Kulvietis G. 167
Лапатин И. 127
Lebedev E. 133
Letunovich Y. 137
Лопухова С. 127
Lukashenko O. 142
Любина Т. 148
Макушенко И. 253
Malinkovsky Yu. 34
Маталыцкий М. 121, 226
Михеев П. 155
Minkevičius S. 161, 167
Мишкой Г.К. 114
Moiseeva S. 219
Morozov E. 22, 142
Науменко В. 172
Назаров А. 148, 179, 186
Нежельская Л. 96
Oukid H. 191
Pagano M. 142
Panfilova T. 271
Паньков А. 172, 196
Печинкин А. 201, 207
Ponomarov V. 213
Rachinskaya M. 82
Разумчик Р. 207
Rusilko T. 230
Satin Ya. 271
Семенова О. 258
Семенова И. 179
Shilova G. 271
Синюгина Ю. 258
Sinyakova I. 219
Статкевич С. 226, 230
Степанова Н. 236
Судыко Е. 186
Сущенко С. 155
Тарамин О. 65, 102
Tikhonenko O. 241
Tsarenkov G. 58
Tsitsiashvili G. 248
Усар И. 253
Вишневский В. 258
Якубович О. 265
Zeifman A. 271
Zorine A. 277
Зуевич В. 88

Научное издание

МАССОВОЕ ОБСЛУЖИВАНИЕ: ПОТОКИ, СИСТЕМЫ, СЕТИ

Материалы международной научной конференции
«Современные вероятностные методы анализа и оптимизации информационно-телекоммуникационных сетей»
г. Минск, 31 января – 3 февраля 2011 г.

Ответственный за выпуск А. Н. Дудин
В авторской редакции
Компьютерная верстка *Г. В. Царенкова*

Подписано в печать 30.12.2010. Формат 60×84/8. Бумага офсетная.
Гарнитура «Times». Ризография.
Усл. печ. л. 32,78. Уч.-изд. л. 20,09. Тираж 80 экз. Зак. № 491.

Отпечатано с оригинал-макета заказчика.
Издатель и полиграфическое исполнение
Государственное учреждение образования
«Республиканский институт высшей школы»
ЛИ № 02330.0133359 от 29.06.2004 г.
220007, Минск, ул. Московская, 15.