



Ссылка на статью:

// Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ.Электрон. журн. 2021, № 1, с.109-117.

Поступила: 11.05.2021

Окончательный вариант: 11.05.2021

© УлГУ

УДК 519.218.5

Математическая модель мультисервисного кол-центра с многоэтапным обслуживанием и дообслуживанием неприоритетных заявок

Савинов Ю. Г.^{*}, Шукин А. Н., Подгорнов М. Д.

^{*}uras@aport.ru

УлГУ, Ульяновск, Россия

В работе развивается семимартингальный (траекторный) подход к математическому описанию и моделированию кол-центров. Рассмотрена модель кол-центра с дообслуживанием неприоритетных заявок и блокировкой вызовов при многоэтапном обслуживании. Показан переход от математической модели к итерационным формулам, по которым проводится имитационное моделирование.

Ключевые слова: мультисервисные кол-центры, многоэтапное обслуживание, блокировка, семимартингальное описание, точечный процесс, компенсатор.

Введение

Современные кол-центры большинства компаний (AVAYA, CISCO и др.) предлагают решения, обеспечивающие интеграцию своих центров обслуживания вызовов (ЦОВ) с Internet. Это означает, что оператор кол-центра должен обрабатывать запросы нескольких типов: звонки, электронные письма, факсимильные запросы, отвечать в чатах. В теории массового обслуживания такие системы принято называть мультисервисными. При этом очень часто обслуживание может состоять в последовательном прохождении нескольких операторов. Системы массового обслуживания (СМО), описывающие такого рода процессы принято называть многофазными, а обслуживание многоэтапным [1-2].

Существует два основных способа математического описания и моделирования мультисервисных кол-центров в терминах СМО: марковский (см., например, работы [3-20]) и семимартингальный (см., например, работы [21-28]). Одним из преимуществ семимартингального (траекторного) описания (в терминах считающих процессов [29]) является легкость перехода от математической модели к итерационным формулам, по которым прово-

дится имитационное моделирование. Поэтому в данной работе, с учетом того, что аналитическое исследование многофазных систем затруднено (они часто теряют марковость даже при простейшем входящем потоке) выбрано семимартингалное описание в терминах точечных (считающих) процессов и их компенсаторов [29].

Чтобы продемонстрировать возможности и простоту использования семимартингалного подхода рассмотрены две типичные ситуации, встречающиеся в современных колл-центрах: дообслуживание неприоритетной заявки и блокировка заявок при переходе к следующему оператору.

Дообслуживание означает, что при поступлении приоритетной заявки (звонка) оператор прекращает обработку неприоритетной (ответ на электронное письмо). После окончания звонка и отсутствия других звонков оператор продолжает обработку прерванной неприоритетной заявки (продолжает отвечать на электронное письмо).

Блокировка означает, что узел обслуживания (обслуживающий прибор) заблокирован, если обслуживание заявки в данном узле завершено, а следующий узел не готов к приему требования по той причине, что в ней не закончено обслуживание. Принимается также, что если узел занят или заблокирован, то очередное входящее требование встает в очередь. Таким образом, в системе могут быть следующие состояния: «узел свободен», «узел занят», «узел заблокирован» [30].

1. Постановка задачи

Рассмотрим мультисервисный колл-центр, в который поступают заявки двух типов (непринципиальное упрощение модели): первый тип – звонки, второй тип - электронные письма. Звонки образуют простейший поток с интенсивностью $\lambda_1 > 0$ (см. рис.1).

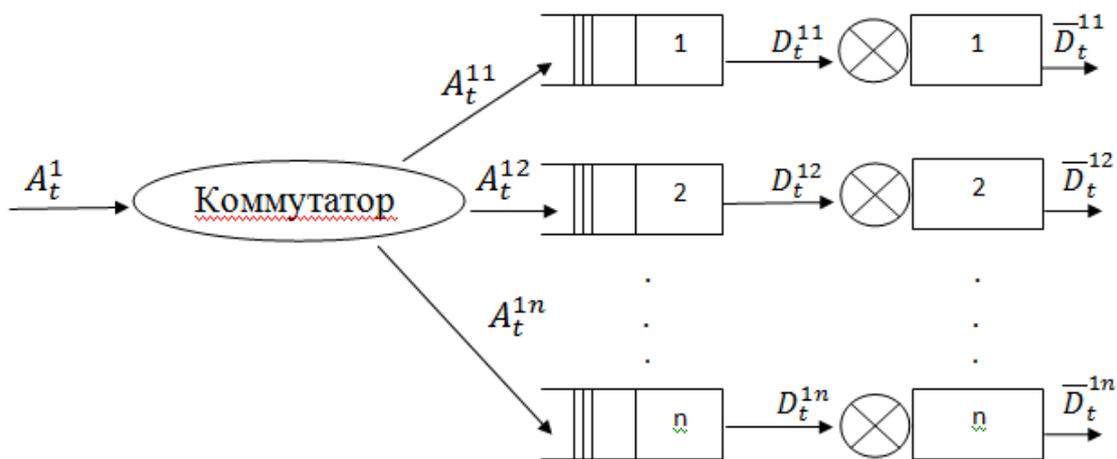


Рис. 1. Схема колл-центра (звонки)

Коммутатор отправляет звонок одному из n операторов первой линии имеющих одинаковую квалификацию со средним временем обслуживания μ_1 . Как принято в реальных

кол-центрах (например, Avaya) в модели вызов передается наиболее свободному оператору, наименее загруженному с начала смены, с учетом коэффициента квалификации. К каждому оператору формируется отдельная неограниченная очередь. Для звонков предусмотрено двухэтапное обслуживание с блокировкой клиента при передаче заявки к следующему оператору (например, техническому специалисту). При двухэтапном обслуживании если первый узел занят (оператор первой линии), то вновь поступившая заявка встает в очередь (размер очереди неограничен). Если второй узел занят (оператор второй линии), то заявка, завершившая обслуживание в первом узле остается в нем (блокирует поступление заявок на обслуживание в первый узел). Операторы второй линии имеют одинаковую квалификацию со средним временем обслуживания μ_{11} .

Поток электронных писем образует простейший поток с интенсивностью $\lambda_2 > 0$ (см. рис.2). Роутер отправляет письмо одному из n операторов имеющих одинаковую квалификацию со средним временем обслуживания μ_2 . Письмо передается наиболее свободному оператору, наименее загруженному с начала смены, с учетом коэффициента квалификации. К каждому оператору формируется отдельная неограниченная очередь.

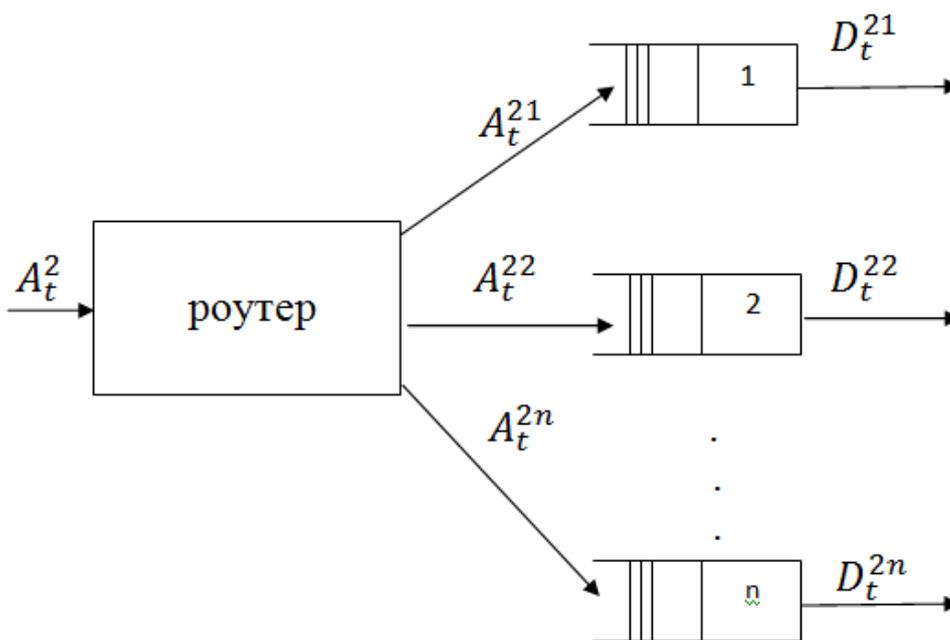


Рис. 2. Схема кол-центра (электронные письма)

Звонки имеют абсолютный приоритет над письмами, но текущее письмо не теряется, а дообслуживается: при поступлении приоритетной заявки (звонка) оператор прекращает обработку неприоритетной (ответа на электронное письмо). После окончания звонка и при отсутствии других звонков оператор продолжает обработку письма.

2. Математическая модель

1. Опишем сначала работу по обработке звонков.

Введем считающие процессы: $A^1 = (A_t^1)_{t \geq 0}$ - число звонков, поступивших в систему за время $t \geq 0$, $A_0^1 = 0$; $D^{1i} = (D_t^{1i})_{t \geq 0}$ - число обслуженных звонков i -тым оператором первой линии ($i = 1, \dots, n$) за время $t \geq 0$, $D_0^{1i} = 0$; $A^{1i} = (A_t^{1i})_{t \geq 0}$ - число звонков, поступивших i -тому оператору первой линии ($i = 1, \dots, n$) за время $t \geq 0$, $A_0^{1i} = 0$; $\bar{D}^{1i} = (\bar{D}_t^{1i})_{t \geq 0}$ - число обслуженных звонков i -тым оператором второй линии ($i = 1, \dots, n$) за время $t \geq 0$, $D_0^{1i} = 0$.

Точечные процессы A^1, D^{1i} ($i = 1, \dots, n$) определяются своими компенсаторами $\tilde{A}^1 = (\tilde{A}_t^1)_{t \geq 0}$ и $\tilde{D}^{1i} = (\tilde{D}_t^{1i})_{t \geq 0}$ в соответствии с разложением Дуба-Мейера для субмартигалов [29]:

$$A_t^1 = \tilde{A}_t^1 + m_t^{A^1}, \quad (1)$$

$$D_t^{1i} = \tilde{D}_t^{1i} + m_t^{D^{1i}}, \quad (2)$$

где \tilde{A}^1 и \tilde{D}^{1i} - неубывающие предсказуемые процессы m^{A^1} и $m^{D^{1i}}$ - мартингалы, $i = 1, \dots, n$.

Для рассматриваемой в данной работе системы $A^1 = (A_t^1)_{t \geq 0}$ - пуассоновский процесс с компенсатором:

$$\tilde{A}_t^1 = \lambda_1 \cdot t, \quad \lambda_1 > 0. \quad (3)$$

Компенсатор для процесса $D^{1i} = (D_t^{1i})_{t \geq 0}$ определяется следующим соотношением:

$$\tilde{D}_t^{1i} = \int_0^t \mu_1 I(\xi_s^{1i} > 0, \bar{\xi}_s^{1i} \leq 1) ds, \quad (4)$$

где $I(\cdot)$ - индикаторная функция, $i = 1, \dots, n$. Здесь ξ_t^{1i} - число клиентов (звонков) у i -ого оператора первой линии в момент времени $t \geq 0$, $\xi_0^{1i} = 0$ (в очереди + обслуживается), $\bar{\xi}_t^{1i}$ - число клиентов (звонков) у i -ого оператора второй линии в момент времени $t \geq 0$ (может быть только 0 или 1 или 2), $\bar{\xi}_0^{1i} = 0$. То есть если i -ый оператор второй линии свободен, то $\bar{\xi}_t^{1i} = 0$, если занят, и i -ый оператор первой линии не заблокирован, то $\bar{\xi}_t^{1i} = 1$; $\bar{\xi}_t^{1i} = 2$, если оператор второй линии занят, и заблокирован оператор первой линии. Таким образом, интенсивность обслуживания заявок у операторов первой линии определяется числом заявок в первой и второй линии, т.е. если второй узел уже занят заявкой, то обслуживаемая заявка из первого узла блокируется в первом узле и ожидает своей очереди.

Для ξ_t^{1i} в момент времени $t \geq 0$ можно написать следующее основное балансовое соотношение

$$\xi_t^{1i} = A_t^{1i} - D_t^{1i}, \quad i = 1, \dots, n. \quad (5)$$

Аналогично для $\bar{\xi}_s^{1i}$

$$\bar{\xi}_t^{1i} = D_t^{1i} - \bar{D}_t^{1i}, \quad i = 1, \dots, n. \quad (6)$$

Чтобы описать логику работы коммутатора (вызов передается наиболее свободному оператору, наименее загруженному с начала смены, с учетом коэффициента квалификации) введем индикаторы:

$$r_t^{11} = I \left(\xi_t^{11} \leq \min_{1 \leq i \leq n} \xi_t^{1i} \right), \quad (7)$$

$$r_t^{1k} = I \left(\xi_t^{1k} \leq \min_{1 \leq i \leq n} \xi_t^{1i} \right) \cdot I(\xi_t^{1k} < \min_{i < k} \xi_t^{1i}), \quad k = 2, \dots, n. \quad (8)$$

То есть, если первый оператор самый незагруженный, то $r_t^{11} = 1$, среди самых незагруженных (если их более одного) выбираем оператора с меньшим номером. Тогда для считающего процесса A^{1i} можно записать:

$$A_t^{1i} = \int_0^t r_s^{1i} dA_s^1, \quad i = 1, \dots, n. \quad (9)$$

Компенсатор для процесса $\bar{D}_t^{1i} = (\bar{D}_t^{1i})_{t \geq 0}$ определяется следующим соотношением:

$$\widetilde{\bar{D}}_t^{1i} = \int_0^t \mu_{11} I(\bar{\xi}_s^{1i} > 0) ds, \quad i = 1, \dots, n, \quad (10)$$

то есть заявки обслуживаются, если второй узел не свободен.

2. Опишем теперь работу по обработке электронных писем.

Введем считающие процессы: $A^2 = (A_t^2)_{t \geq 0}$ - число писем, поступивших в систему за время $t \geq 0$, $A_0^2 = 0$; $D^{2i} = (D_t^{2i})_{t \geq 0}$ - число обслуженных писем i -тым оператором первой линии ($i = 1, \dots, n$) за время $t \geq 0$, $D_0^{2i} = 0$; $A^{2i} = (A_t^{2i})_{t \geq 0}$ - число писем, поступивших i -тому оператору первой линии ($i = 1, \dots, n$) за время $t \geq 0$, $A_0^{2i} = 0$. Операторы второй линии в обработке писем не участвуют.

Точечные процессы A^2, D^{2i} ($i = 1, \dots, n$) определяются своими компенсаторами $\tilde{A}^2 = (\tilde{A}_t^2)_{t \geq 0}$ и $\tilde{D}^{2i} = (\tilde{D}_t^{2i})_{t \geq 0}$ в соответствии с разложением Дуба-Мейера для субмартингалов [29]:

$$A_t^2 = \tilde{A}_t^2 + m_t^{A^2}, \quad (11)$$

$$D_t^{2i} = \tilde{D}_t^{2i} + m_t^{D^{2i}}, \quad (12)$$

где \tilde{A}^2 и \tilde{D}^{2i} - неубывающие предсказуемые процессы m^{A^2} и $m^{D^{2i}}$ - мартингалы, $i = 1, \dots, n$.

Для рассматриваемой в данной работе системы $A^2 = (A_t^2)_{t \geq 0}$ - пуассоновский процесс с компенсатором:

$$\widetilde{A}_t^2 = \lambda_2 \cdot t, \lambda_2 > 0. \quad (13)$$

Компенсатор для процесса $D^{2i} = (D_t^{2i})_{t \geq 0}$ определяется следующим соотношением:

$$\widetilde{D}_t^{2i} = \int_0^t \mu_2 I(\xi_s^{1i} = 0, \xi_s^{2i} > 0) ds, \quad (14)$$

где $I(\cdot)$ – индикаторная функция, $i = 1, \dots, n$. Здесь ξ_t^{2i} – число писем у i -ого оператора первой линии (в очереди + обслуживается) в момент времени $t \geq 0$, $\xi_0^{1i} = 0$. Соответственно индикатор $I(\xi_s^{1i} = 0)$ проверяет отсутствие других звонков у оператора и продолжает обработку письма если есть что продолжать, то есть когда $\xi_s^{2i} > 0$.

Для ξ_t^{2i} в момент времени $t \geq 0$ можно написать следующее основное балансовое соотношение

$$\xi_t^{2i} = A_t^{2i} - D_t^{2i}, \quad i = 1, \dots, n. \quad (15)$$

Логике работы роутера (письмо передается наиболее свободному оператору, наименее загруженному с начала смены, с учетом коэффициента квалификации) также можно описать с помощью индикаторов:

$$r_s^{21} = I \left(\xi_s^{21} \leq \min_{1 \leq i \leq n} \xi_s^{2i} \right), \quad (16)$$

$$r_s^{2k} = I \left(\xi_s^{2k} \leq \min_{1 \leq i \leq n} \xi_s^{2i} \right) \cdot I \left(\xi_s^{2k} < \min_{i > k} \xi_s^{2i} \right), \quad k = 2, \dots, n. \quad (17)$$

То есть, если первый оператор самый незагруженный, то $r_t^{21} = 1$, среди самых незагруженных (если их более одного) выбираем оператора с большим номером (противоположная стратегия в сравнении со звонками). Тогда для считающего процесса A^{2i} можно записать:

$$A_t^{2i} = \int_0^t r_s^{2i} dA_s^2, \quad i = 1, \dots, n. \quad (18)$$

3. Итерационные формулы для компьютерного моделирования

Выведем формулы, позволяющие произвести имитационное моделирование СМО. На стохастическом базисе $B = (\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t \geq 0}, P)$ из формул (1)-(18) можно получить следующие соотношения:

$$P\{A_{t+\Delta}^1 - A_t^1 = 1 | \mathcal{F}_t\} = \lambda_1 \cdot \Delta + o(\Delta), \quad (19)$$

$$P\{A_{t+\Delta}^2 - A_t^2 = 1 | \mathcal{F}_t\} = \lambda_2 \cdot \Delta + o(\Delta), \quad (20)$$

$$P\{D_{t+\Delta}^{1i} - D_t^{1i} = 1 | \mathcal{F}_t\} = \mu_1 \cdot I(\xi_t^{1i} > 0, \bar{\xi}_t^{1i} \leq 1) \cdot \Delta + o(\Delta), \quad (21)$$

$$P\{D_{t+\Delta}^{2i} - D_t^{2i} = 1 | \mathcal{F}_t\} = \mu_2 \cdot I(\xi_t^{2i} = 0, \xi_t^{2i} > 0) \cdot \Delta + o(\Delta), \quad (22)$$

$$P\{\bar{D}_{t+\Delta}^{1i} - \bar{D}_t^{1i} = 1 | \mathcal{F}_t\} = \mu_{11} \cdot I(\bar{\xi}_t^{1i} > 0) \cdot \Delta + o(\Delta), \quad (23)$$

Формулы (19)-(23) позволяют, основываясь на понятии геометрической вероятности, провести имитационное моделирование. А именно, введя дискретизацию (шаг по времени) Δ из условия $\lambda_i \cdot \Delta \ll 1, \mu_i \cdot \Delta \ll 1, \mu_{11} \cdot \Delta \ll 1, i = 1, 2$ получим следующие итерационные формулы (для вычисления последующих значений через предыдущие):

$$A_{t+\Delta}^1 = A_t^1 + \delta(\lambda_1), \quad (24)$$

$$A_{t+\Delta}^2 = A_t^2 + \delta(\lambda_2), \quad (25)$$

$$D_{t+\Delta}^{1i} = D_t^{1i} + \delta(I(\xi_t^{1i} > 0, \bar{\xi}_t^{1i} \leq 1) \cdot \mu_1), \quad (26)$$

$$D_{t+\Delta}^{2i} = D_t^{2i} + \delta(I(\xi_t^{2i} = 0, \xi_t^{2i} > 0) \cdot \mu_2), \quad (27)$$

$$\bar{D}_{t+\Delta}^{1i} = \bar{D}_t^{1i} + \delta\left(\mu_{11} \cdot I(\bar{\xi}_t^{1i} > 0)\right), \quad (28)$$

где $\delta(\gamma) = \begin{cases} 1, & \text{с вероятностью } \gamma \cdot \Delta, \\ 0, & \text{с вероятностью } 1 - \gamma \cdot \Delta. \end{cases}$

$$A_{t+\Delta}^{1i} = A_t^{1i} + r_s^{1i} \cdot I(A_{t+\Delta}^1 - A_t^1 = 1), \quad (29)$$

$$A_{t+\Delta}^{2i} = A_t^{2i} + r_s^{2i} \cdot I(A_{t+\Delta}^2 - A_t^2 = 1). \quad (30)$$

Заключение

В результате выполнения данной работы была построена математическая модель кол-центра с дообслуживанием неприоритетных заявок и блокировкой вызовов при многоэтапном обслуживании в семимартингальных терминах. Продемонстрированы возможности и простота использования траекторного подхода при математическом моделировании кол-центров. Показан переход от математической модели к итерационным формулам, по которым может проводиться имитационное моделирование.

Список литературы

1. Матвеев В.Ф. Многоэтапные системы обслуживания / в сборнике *Вычислительные методы и программирование*. М.: Изд-во МГУ. 1972, том 18, с. 55-79.
2. Бутов А.А., Савинов Ю.Г. *Теория массового обслуживания: Учебно-методическое пособие*. Ульяновск: УлГУ, 2007.
3. Грачев В.В., Моисеев А.Н., Назаров А.А., Ямпольский В.З. Многофазная модель массового обслуживания системы распределенной обработки данных // *Доклады ТУСУРа*. 2012, № 2 (26), часть 2, 251с.
4. Ануфриев Д.П., Холодов А.Ю. *Статистический анализ имитационных экспериментов модели системы массового обслуживания с накопителем и интервальной за-*

- держкой начала обслуживания. Астраханский инженерно-строительный институт (ГАОУ АО ВПО «АИСИ»), 2014.
5. Гаджиев А.Г., Мамедов Т.Ш. Циклические системы с задержкой обслуживания // *Доклады Академии наук*. 2009, т.26, № 1, с.15-19.
 6. Глушак Е.В. Исследование и разработка математических моделей распределенных центров обслуживания вызовов: дис. ... канд. техн. наук: 05.12.13. Самара, 2014. 134 с.
 7. Глушак, Е.В. Исследование моделей распределенного центра обслуживания вызовов / Е.В. Глушак, А.В. Росляков // *Тез. докл. на XII Международной научно-технической конференции «Проблемы техники и технологии телекоммуникаций»*, Казань, 2011. С. 185-186.
 8. Palm C. Intensitat schwankungen im fernsprechverkehr // *Ericsson technics*. 1943, № 44, p. 1-189.
 9. Риордан Д. *Вероятностные системы обслуживания*. Пер. с англ. М.: Связь, 1966.
 10. Baccelli F., Hebuterne G. On queues with impatient customers // *Performance, North-Holland publishing company*. 1981, p. 159-179.
 11. Voxma O.J., de Waal P.R. Multiserver queues with impatient customers // *ITC*. 1994, № 14, p. 743-756.
 12. Haugen R.B., Skogan E. Queueing systems with stochastic time out // *IEEE Transactions on Communications*. 1980, № 28, p.1984-1989.
 13. Brandt A., Brandt M. On the $M(n)/M(s)/s$ queue with impatient calls // *Performance evaluation*. 1999, № 35, p.1-18.
 14. Brandt A., Brandt M. Asymptotic results and a Markovian approximation for the $M(n)/M(s)/s+GI$ system // *Queueing systems: theory and applications (QUESTA)*. 2002, № 41, p. 73-94.
 15. Palm C. Research on telephone traffic carried by full availability groups // *Tele*. 1957. № 1, p. 107.
 16. Garnett O., Mandelbaum A., Reiman M.I. Designing a call center with impatient customers // *Manufacturing and Service Operations Management*. 2002, № 4, p. 208-227.
 17. Aguir S., Karaesmen F., Aksin O.Z., Dallery Y. On the interaction between retrials and sizing of call centers // *Working paper, Koc University, Department of Industrial Engineering*. 2004, p. 434.
 18. Росляков, А.В., Цыганков Н.И. Анализ моделей распределенных центров обслуживания вызовов // *Электросвязь*, № 8. 2005, с. 22-25.
 19. Зарипова Э.Р. Математическая модель центра обслуживания вызовов с двумя типами абонентов // *Discrete and Continuous Models and Applied Computational Science*. 2010, № 4, с. 76-82.
 20. Ваняшин С.В. Исследование и разработка моделей мультисервисного центра обслуживания вызовов: дис. ... канд. тех. наук: 05.12.13. М.: 2006. 157 с.

21. Бутов А.А., Галимов Л.А. Оптимальное управление интенсивностью входящего потока многоканальной СМО с роутером при эпизодически наблюдаемой длине очереди на приборах // *Современные проблемы науки и образования*. 2015, № 2, с. 758.
22. Бутов А.А., Галимов Л.А. Стохастическая имитационная модель оценки резерва произошедших, но не заявленных страховых убытков в терминах СМО // *Фундаментальные исследования*. 2016, № 8-2, с. 234-238.
23. Бутов А.А., Галимов Л.А. Оптимальное управление распределением заявок в многоканальной системе массового обслуживания с входящим пуассоновским потоком поступления заявок и экспоненциальным временем обслуживания // *Естественные и технические науки*. 2014, № 9-10 (77), с. 244-247.
24. Столяров И.А., Табакова Е.Д., Савинов Ю.Г., Столяров И.А. Семимартингальная модель многоканальной СМО с ограниченным временем ожидания // *Прикладная математика и информатика: современные исследования в области естественных и технических наук: материалы IV научно-практической всероссийской конференции (школы-семинара) молодых ученых: 23-25 апреля 2018 г. В двух частях.* – Тольятти: Издатель Качалин Александр Васильевич, 2018, Часть 1, С. 502-506.
25. Столяров И.А., Савинов Ю.Г. Семимартингальная модель СМО с динамическим приоритетом // *Прикладная математика и информатика: современные исследования в области естественных и технических наук: материалы III научно-практической всероссийской конференции (школы-семинара) молодых ученых: 24-25 апреля 2017 г.* – Тольятти: Издатель Качалин Александр Васильевич, 2017. – стр. 553-557.
26. Савинов Ю.Г., Исмаилова М.В., Рослов М.Э. Траекторные методы моделирования многофазных СМО // *Ученые записки УлГУ. Серия: Математика и информационные технологии*. 2019, № 1, с. 85-91.
27. Савинов Ю.Г., Рослов М.Э., Куманина Я.А. Стратегия адаптивного разделения ресурсов в мультисервисных сетях // *Ученые записки УлГУ. Серия: Математика и информационные технологии*. 2020, № 1, с. 95-102.
28. Савинов Ю.Г., Тихоненко А.А., Пронин В.И., Щукин А.Н. Семимартингальная модель СМО с произвольным временем ожидания «нетерпеливых» заявок // *Ученые записки УлГУ. Серия: Математика и информационные технологии*. 2019, № 2, с. 81-88.
29. Бутов А.А., Раводин К.О. *Теория случайных процессов: учебное пособие*. Ульяновск: УлГУ, 2009.
30. Афонин В.В., Федосин С.А. *Моделирование систем*. Интуит, 2016. 2-е изд.