

ЧИСЛЕННЫЙ АНАЛИЗ ДИФФУЗИОННОЙ АППРОКСИМАЦИИ ОДНОКАНАЛЬНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С УПРАВЛЯЕМЫМИ ПАРАМЕТРАМИ ЗАГРУЗКИ И ОБРАБОТКИ

К.В. Семенихин, Ю. С. Иванова

Работа посвящена численному анализу качества диффузионной аппроксимации управляемой одноканальной системы массового обслуживания в загруженном состоянии. Управление системой задается вероятностью отказа и интенсивностью обработки. Цель исследования - нахождение условий применимости диффузионной аппроксимации к модели управляемого марковского процесса, описывающего состояние системы. Качество аппроксимации характеризуется величиной равномерной метрики, определяющей расстояние между стационарными распределениями в исходной и приближенной модели, а также числовыми показателями, такими как средняя загрузка, среднее время ожидания, доля отклоненных заявок и объем затраченных ресурсов. Аппроксимация оказывается тем точнее, чем с большей вероятностью диффузионный процесс попадает в интервал доступа, т.е. интервал, в котором вероятность отказа меняется от нуля до единицы.

The paper is devoted to numerical analysis of diffusion approximation accuracy for overloaded controlled single-channel queuing systems. Active queue management is defined by two controlled parameters: rejection probability and service rate. The aim of the study is to find conditions for the applicability of diffusion approximation to the queuing model governed by a continuous-time controlled Markov chain. The accuracy of the approximation is characterized by the uniform metric that specifies the distance between steady-state distributions in the original and approximate models, as well as by several indicators such as average load, average waiting time, the percentage of rejected packets and the amount of resources spent. The approximation is the more accurate, the more likely the diffusion process falls into the access interval - the interval in which the rejection probability varies from zero to one.

КЛЮЧЕВЫЕ СЛОВА

Система массового обслуживания, управление доступом, управление загрузкой, дискретный марковский процесс с непрерывным временем, диффузионная аппроксимация, модель линейной диффузии.

ДЛЯ ЦИТАТЫ

К.В. Семенихин, Ю.С. Иванова. Численный анализ диффузионной аппроксимации одноканальной системы массового обслуживания с управляемыми параметрами загрузки и обработки // Моделирование и анализ данных. 2019. №2. С.21-30.

K.V. Semenikhin, Yu.S. Ivanova. Numerical analysis of diffusion approximation for a single-channel queuing system with controlled access and service rates. Modelirovaniye i analiz dannykh=Modelling and data analysis (Russia). 2019, no.2, pp.21-30.

1. ВВЕДЕНИЕ

Современные системы передачи и обработки информации по своей сути являются системами массового обслуживания (СМО), в которых входной поток состоит из однотипных пакетов — небольших порций данных, предусмотренных сетевым протоколом [1,2]. Традиционные методы анализа и оптимизации СМО основаны на использовании марковских процессов с конечным множеством состояний [3,4].

Однако при значительном объеме трафика практическое применение данной методологии становится затруднительным из-за большого числа уравнений, описывающих стационарные вероятности состояний соответствующей СМО. Один из эффективных способов решения проблемы большой размерности, возникающей при оптимизации дискретных марковских процессов, является метод диффузионной аппроксимации [5,6].

Аналогично [7-9] в данной работе рассматривается одноканальная СМО, имеющая конечную очередь, принимающая на вход простейший поток и предусматривающая возможность управления характеристиками доступа и загрузки: вероятность отказа и интенсивность обработки. Однако в отличие от указанных работ далее система будет изучаться в условиях сильной загрузки. Кроме того, управление осуществляется за счет выбора параметров двух линейных функций, определяющих зависимость вероятности отказа и интенсивности обработки от числа заявок в системе. Обе функции отражают естественное требование роста числа отброшенных пакетов и скорости обработки при увеличении загрузки системы. Целью работы является нахождение условий применимости диффузионной аппроксимации к модели управляемого марковского процесса, описывающего число заявок в системе. В работе предлагается диффузионная аппроксимация, основанная на модели стохастического дифференциального уравнения с линейным сносом и диффузией [10]. Устанавливается, что в зависимости от соотношения параметров системы стационарное распределение аппроксимирующего процесса является трехпараметрическим гамма-распределением или гауссовским.

Если назвать интервалом доступа тот промежуток, в котором вероятность отказа как функция состояния меняется в интервале $(0,1)$. Тогда основной результат работы можно сформулировать следующим образом: качество аппроксимации зависит от вероятности попадания диффузионного процесса в интервал доступа. Поэтому если вероятность отказа априори отделена снизу от нуля и сверху от единицы, то диффузионная аппроксимация вполне приемлема для приближенного описания и последующей оптимизации управляемой СМО с интенсивным входным потоком.

2. ОПИСАНИЕ МОДЕЛИ

Рассмотрим систему массового обслуживания с конечной очередью и одним прибором, обслуживающим заявку за экспоненциальное время $E(\mu)$. Максимальное число заявок в системе равно N . Изначально поступающий поток заявок является простейшим интенсивности $\alpha > 0$. Однако этот поток прореживается с помощью процедуры геометрического просеивания, в результате чего входной поток, оставаясь простейшим, имеет интенсивность $\alpha(1-r)$, где $r \in [0,1]$ — вероятность отклонения приходящей заявки. Эволюцию числа заявок в системе $M_{\alpha(1-r)}|M_{\mu}|1|N-1$ можно описывать с помощью однородного марковского процесса $\{X(t), t \geq 0\}$ с множеством состояний $E = \{0, 1, \dots, N\}$. Указанный марковский процесс относится к классу процессов рождения и гибели с интенсивностями перехода:

$$a) \lambda_{x,x+1} = \alpha(1-r) \text{ для } x = 0, 1, \dots, N-1;$$

$$b) \lambda_{x-1,x} = \mu \text{ для } x = 1, 2, \dots, N;$$

Предположим, что параметрами входного потока и обслуживающего прибора можно управлять. Пусть вероятность отклонения заявки $r(x)$ и интенсивность обработки $\mu(x)$ являются функциями состояния $x \in E$, а именно:

$$(1) \quad r(x) = \text{proj}_{[0,1]}(r_0 + r_1 x), \quad \mu(x) = \mu_0 + \mu_1 x$$

где $\text{proj}_{[0,1]}(z)$ обозначает проекцию точки z на отрезок $[0,1]$. Роль параметров управления здесь выполняют числа $r_0 < 1, r_1 > 0, \mu_0 > 0, \mu_1 \geq 0$.

В этом случае процесс $X(t)$, равный числу заявок в системе, остается однородным марковским, но его интенсивности переходов теперь зависят от состояния:

$$a) \lambda_{x,x+1} = \alpha(1-r(x)) \text{ для } x = 0, 1, \dots, N-1;$$

$$b) \lambda_{x-1,x} = \mu(x) \text{ для } x = 1, 2, \dots, N;$$

Полученную управляемую СМО будем обозначать $M_{\alpha(1-r(x))} | M_{\mu(x)} | 1 | N - 1$.

По построению вероятность отказа $r(x)$ растет с увеличением загрузки системы, что подобно алгоритму RED (Random Early Detection), предназначенному для отклонения случайной доли входящего трафика для сигнализации передающей стороне о возникновении коллизий [12].

Для функции $r(x)$ в дальнейшем будет использоваться также и другая параметризация:

$$r(x) = \begin{cases} 0, & x \leq x_0; \\ (x - x_0)/(x_1 - x_0), & x_0 \leq x \leq x_1; \\ 1, & x \geq x_1, \end{cases}$$

где x_0 обозначает состояние, вплоть до которого все заявки принимаются, а x_1 — состояние, начиная с которого все входящие заявки отклоняются.

Основную роль в дальнейшем играет промежуток (x_0, x_1) , который будем называть *интервалом доступа*, так как именно на нем осуществляется управление доступом к ресурсам СМО.

Кроме того, все входящие заявки принимаются при $X(t) \geq x_0$ и, наоборот, отклоняются при $X(t) \leq x_1$. Поэтому $[0, x_0]$ — *интервал приема* и соответственно $[x_1, N]$ — *интервал блокировки*.

Интенсивность обработки также является возрастающей функцией, что соответствует идеологии семейства протоколов TCP/AQM, увеличивающих скорость приема при росте запросов [13]. При этом $\mu_{min} = \mu_0$,

$\mu_{max} = \mu_0 + \mu_1 N$ — это соответственно минимальное и максимальное значения интенсивности обработки. Характерный вид зависимостей $r(x)$, $\mu(x)$ от состояния системы X указан на рис. 1.

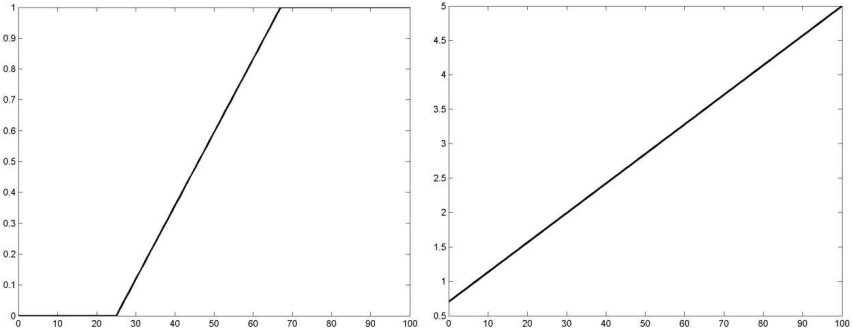


Рис. 1. Вероятность отказа $r(x)$ (слева) и интенсивности обработки $\mu(x)$ (справа).

3. МЕТОД ДИФFUЗИОННОЙ АППРОКСИМАЦИИ

В качестве диффузионной аппроксимации исходного дискретного марковского процесса $X(t)$, будем рассматривать процесс $\xi(t)$ удовлетворяющий стохастическому дифференциальному уравнению

$$(2) \quad d\xi(t) = A(\xi(t))dt + B(\xi(t))dW(t),$$

где $A(x), B(x)$ — специально подобранные коэффициенты сноса и диффузии, а $W(t)$ — стандартный винеровский процесс, не зависящий от начального состояния $\xi(0)$.

Поскольку $\xi(t)$ является в то же время однородным марковским процессом, коэффициенты сноса и диффузии могут быть заданы через его переходную вероятность

$$(3) \quad A(x) = \lim_{h \downarrow 0} \frac{1}{h} \int_{-\infty}^{\infty} (y - x) P(x, h, dy), \quad B(x) = \lim_{h \downarrow 0} \frac{1}{h} \int_{-\infty}^{\infty} (y - x)^2 P(x, h, dy),$$

если выполнено условие

$$\lim_{h \downarrow 0} \frac{1}{h} \int_{|y-x|>\delta}^{\infty} (1 + (y - x)^2) P(x, h, dy) = 0 \quad \forall \delta > 0,$$

где $P(x, h, dy) = P\{\xi(t+h) \in E | \xi(t) = x\}$ обозначает переходную вероятность процесса $\xi(t)$.

Для нахождения коэффициентов $A(x), B(x)$ введем допущение о том, что переходную вероятность $P(x, h, \cdot)$ можно достаточно точно приблизить условным распределением $Law\{X(t+h) | X(t) = x\}$ исходного дискретного процесса $X(t)$. С учетом этого в (3) ин-

теграл заменяется на сумму по $y \in E$ с весами в виде переходных вероятностей $p_{x,y}(h) = P\{X(t+h) = y | X(t) = x\}$.

Поскольку пределы в (3) представляют собой операцию дифференцирования по h , получаем

$$A(x) = \sum_{y \in E} (y - x) \lambda_{x,y} = \lambda_{x,x+1} - \lambda_{x,x-1},$$

$$B^2(x) = \sum_{y \in E} (y - x)^2 \lambda_{x,y} = \lambda_{x,x+1} + \lambda_{x,x-1}, \text{ откуда}$$

$$(4) \quad A(x) = \alpha(1 - r(x)) - \mu(x), \quad B^2(x) = \alpha(1 - r(x)) + \mu(x).$$

Сделаем еще одно важное упрощение. Будем считать, что в выражении (4) вероятность отказа $r(x)$ определяется линейной зависимостью $r_0 + r_1 x$ без использования операции проектирования на отрезок (1). Тогда система (2) определяет модель диффузионного процесса с линейным сносом и диффузией

$$(5) \quad d\xi(t) = (a_0 + a_1 \xi(t)) dt + \sqrt{b_0 + b_1 \xi(t)} dW(t),$$

где коэффициенты выражаются через функции доступа и загрузки $r(x), \mu(x)$:

$$a_0 = \alpha(1 - r_0) - \mu_0, \quad a_1 = -\alpha r_1 - \mu_1$$

$$b_0 = \alpha(1 - r_0) + \mu_0, \quad b_1 = -\alpha r_1 + \mu_1$$

Итак, цель данной работы — исследовать точность диффузионной аппроксимации (5) для приближенного описания управляемой СМО $M_{\alpha(1-r)} | M_{\mu} | 1 | N - 1$, рассматриваемой в стационарном режиме. Точность диффузионной аппроксимации будем определять с помощью равномерной метрики между функциями распределения исходного дискретного процесса и соответствующего диффузионного процесса:

$$(6) \quad \delta = \sup_{x \in \mathbb{R}} |P\{X \leq x\} - P\{\xi \leq x\}|.$$

Здесь и далее X и ξ — обозначения случайных величин, имеющих стационарное распределение процессов $X(t)$ и $\xi(t)$ соответственно. Для визуального сравнения этих двух распределений будут использоваться графики двух функций

$$(7) \quad p(x) = P\{X = x\}, \quad p_{\xi}(x) = \lim_{x \in \Delta: |\Delta| \rightarrow 0} \frac{P\{\xi \in \Delta\}}{|\Delta|},$$

где $p(x)$ — вероятность состояния дискретного процесса, $p_{\xi}(x)$ — одномерная плотность диффузионного процесса.

Важной характеристикой СМО является средняя загрузка, т.е. среднее число заявок в системе $L = EX$. В качестве аппроксимации будет использоваться $L_{\xi} = E\xi$. Кроме того, для описания качества функционирования СМО вводятся три показателя: среднее время ожидания W ; доля отклоненных заявок R ; объем затраченных ресурсов M . Их аппроксима-

ции, вычисленные с помощью диффузионного процесса, обозначаются соответственно W_ξ , R_ξ и M_ξ .

4. ХАРАКТЕРИСТИКИ ДИФФУЗИОННОГО ПРОЦЕССА

В силу $a_1 < 0$ уравнение (5) определяет отрицательную обратную связь. Поэтому процесс $\xi(t)$ — эргодический [10, 18], т.е. имеет место сходимость по распределению

$$(8) \quad \xi(t) \xrightarrow{d} \bar{\xi},$$

где распределение случайной величины $\bar{\xi}$ не зависит от выбора начального распределения $Law\{\xi(0)\}$. Распределение величины $\bar{\xi}$ будем называть *стационарным распределением диффузионного процесса*. Его конструкция выражается через гамма-распределение $\Gamma(a, \lambda)$, имеющее плотность

$$g(x; a, \lambda) = \frac{\lambda}{\Gamma(a)} (\lambda x)^{a-1} e^{-\lambda x}, x > 0$$

где a и λ — положительные параметры, $\Gamma(a)$ — гамма-функция Эйлера.

Если обозначить

$$(9) \quad \eta \sim \Gamma(a, \lambda), \quad a = \frac{2(a_0 b_1 - a_1 b_0)}{b_1^2}, \quad \lambda = \frac{2|a_1|}{|b_1|}, \quad c = -\frac{b_0}{b_1}$$

то параметры a и λ будут заведомо положительными. Поэтому указанная случайная величина η определена корректно. С ее помощью можно описать стационарное распределение диффузионного процесса.

Для этого укажем конструкцию случайной величины $\bar{\xi}$:

а) если $b_1 < 0$, то

$$(10) \quad \bar{\xi} = c - \eta, \quad \bar{\xi} \in (-\infty, c), \quad c > 0;$$

б) если $b_1 > 0$, то

$$(11) \quad \bar{\xi} = c + \eta, \quad \bar{\xi} \in (c, +\infty), \quad c < 0;$$

с) если же $b_1 = 0$, то $\bar{\xi}$ — нормальная случайная величина с параметрами

$$(12) \quad E\bar{\xi} = \frac{a_0}{|a_1|}, \quad D\bar{\xi} = \frac{b_0^2}{(2|a_1|)}.$$

Итак, в зависимости от знака коэффициента b_1 носитель стационарного распределения диффузионного процесса будет приходиться на левую или правую полуось. Условие $b_1 < 0$ или, что то же самое, $\alpha r_1 > \mu_1$ означает, что увеличение числа отказов происходит быстрее, чем наращивается интенсивность обработки.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ЭКСПЕРИМЕНТА

Численный анализ качества предложенной аппроксимации проведем для двух вариантов управления рассматриваемой системы массового обслуживания. В первом случае управление будет определяться только интервалом доступа (x_0, x_1) при фиксированных параметрах обслуживания μ_0, μ_1 . Во втором варианте модели управления для заданного интервала доступа (x_0, x_1) изучим влияние на качество аппроксимации выбора параметров обрабатываемого устройства μ_0, μ_1 .

Итак, рассмотрим первый вариант, в котором параметры системы выбраны следующим образом:

$$N = 100, 10 < x_0 < 50, 80 < x_1 < 100, \mu_0 = 1.1, \mu_1 = 0$$

Показатель δ , определяющий расстояние (6) между стационарными распределениями исходного дискретного процесса $X(t)$ и его диффузионной аппроксимацией $\xi(t)$, лежит в диапазоне $\delta_{\min} < \delta < \delta_{\max}$, где $\delta_{\min} = 0.025$, $\delta_{\max} = 0.048$. Наименьшая погрешность аппроксимации δ_{\min} достигается при выборе максимально возможного интервала доступа $(x_0, x_1) = (10, 100)$. Наихудший случай с точки зрения величины δ_{\max} реализуется для интервала $(x_0, x_1) = (50, 80)$, который является минимально возможным на рассматриваемом множестве параметров системы. Полученный результат можно объяснить тем, что при наиболее широком интервале доступа вероятность отказа $r(t)$ наилучшим образом аппроксимируется линейной зависимостью, которая используется в модели диффузионного процесса (5). И наоборот, узкий интервал доступ может привести к тому, что модель процесса с линейной диффузией (5) перестанет служить адекватным описанием управляемой системы массового обслуживания. Тем не менее для рассматриваемого набора параметров даже наихудший уровень аппроксимации остается приемлемым. Это подтверждается сравнением значений средней загрузки: $L = L_\xi = 34.29$ при $\delta = \delta_{\min}$ и $L = 57.91 L_\xi = 58.11$ при $\delta = \delta_{\max}$. Кроме того, в справедливости данного вывода можно убедиться непосредственно, визуальнo сравнив стационарные вероятности дискретного процесса $p(x)$ и стационарную плотность диффузионного процесса $p_\xi(x)$ (см. рис. 2).

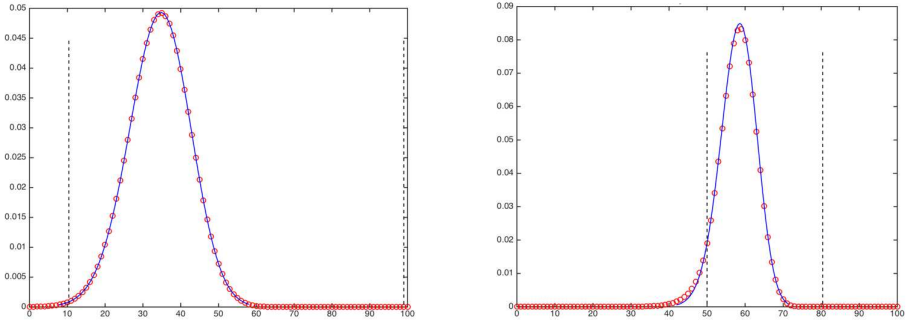


Рис. 2. Вероятности состояний дискретного процесса $p(x)$ (кружочки) и плотность диффузионного процесса $p_{\xi}(x)$ (сплошная кривая) вместе с границами интервала доступа (x_0, x_1) (штриховые линии) при фиксированных параметрах загрузки: слева — случай $\delta = \delta_{\min}$, справа случай $\delta = \delta_{\max}$.

Дальнейшие результаты сравнения при различном объеме очереди N приведены в следующей таблице.

N	δ_{\max}	L	L_{ξ}	W	W_{ξ}	R	R_{ξ}	M	M_{ξ}
50	0.0852	28.5406	29.0533	29.3120	26.4121	0.3510	0.2667	1.1	1.1
100	0.0485	57.9109	58.1067	50.1537	44.5091	0.3495	0.2667	1.1	1.1
500	0.0189	290.532	290.533	250.423	222.545	0.3483	0.2667	1.1	1.1
1000	0.0134	581.0667	581.0667	500.730	445.0909	0.3482	0.2667	1.1	1.1
2000	0.0094	1162.13	1162.13	1001.34	890.1818	0.3481	0.2667	1.1	1.1

Теперь рассмотрим второй вариант управляемой модели, в которой выбран фиксированный интервал доступа с границами $x_0 = 20$ и $x_1 = 80$ при $N=100$. Множество значений параметров обслуживания μ_0, μ_1 описывается следующими условиями:

$$0.2 < \mu_0 < 1.8, 0 < \mu_1 < 0.025.$$

Второе ограничение вытекает из предположения $b_1 < 0$ или, что то же самое, $\mu_1 < \alpha/(x_1 - x_0)$, которое также означает, что система находится в загруженном состоянии. Случай $b_1 > 0$ соответствует малой нагрузке системы, при которой носитель диффузионного процесса располагается за пределами исходного интервала состояний. На рис.3 изображены функции $p(x)$ и $p_{\xi}(x)$, определяющие стационарные распределения в исходной и приближенной модели.

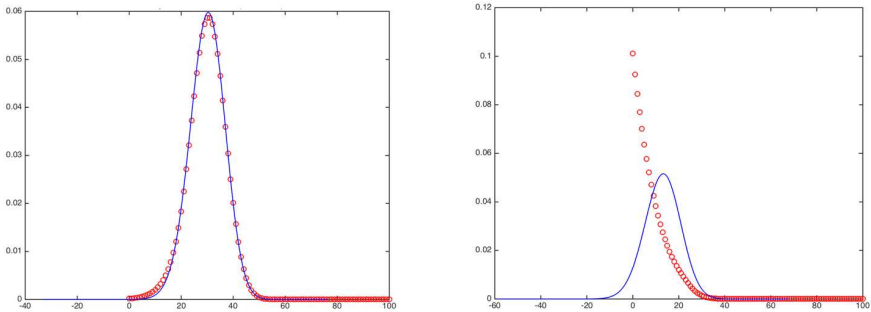


Рис. 3. Вероятности состояний дискретного процесса $p(x)$ (кружочки) и плотность диффузионного процесса $p_{\xi}(x)$ (сплошная кривая) при фиксированном интервале доступа: слева - случай $\delta = 0.032$; справа - случай $\delta = 0.034$.

На левом графике рис.3 представлен случай, в котором метрика δ принимает значение 0.032, т.е. качество диффузионной аппроксимации оказывается вполне приемлемым, хотя носитель диффузионного процесса немного выходит за границы заданного интервала (x_0, x_1) .

Для рассматриваемого множества параметров качество диффузионной аппроксимации может оказаться неприемлемым. Этот случай изображен на правом графике рис.3. Здесь расстояние между исходным распределением и его аппроксимации достигает $\delta = 0.034$. Как видно, носитель стационарного распределения диффузионного процесса находится вне большей части интервала доступа. Именно это обстоятельство приводит к большой погрешности аппроксимации.

6. ЗАКЛЮЧЕНИЕ

В работе рассмотрена модель одноканальной системы массового обслуживания с управляемыми доступом и загрузкой. Вероятность отклонения и интенсивность обслуживания предполагались линейно зависящими от состояния системы. Предложена диффузионная аппроксимация на базе стохастического дифференциального уравнения с линейными коэффициентами сноса и диффузии. Для трех показателей работы системы (среднее время пребывания, доля отказов, средний объем ресурсов) определена зависимость от управляющих параметров. Численный анализ показал, что качество диффузионной аппроксимации стационарного распределения дискретного марковского процесса зависит от вероятности попадания диффузионного процесса в интервал доступа. Если эта вероятность близка к единице (это характерно для ситуации сильной загрузки), то модель диффузионного процесса можно использовать для приближенного описания управляемой одноканальной системы массового обслуживания.

ЛИТЕРАТУРА

1. Степанов С. Н. Основы телетрафика мультисервисных сетей.// М.: Эко-Трендз, 2010.
2. Boucherie R. J., van Dijk N. M. Queueing Networks: A Fundamental Approach.// New York: Springer, 2011.
3. Дынкин Е. Б., Юшкевич А. А. Управляемые марковские процессы и их приложения. М.: Наука, 1975.
4. Stidham S., Weber R. A survey of Markov decision models for control of networks of queues.// Queueing Systems, 1993, Vol. 13, no. 1., Pp. 291-314.
5. Anisimov V. V. Switching processes in queueing models.// Hoboken: J. Wiley&Sons, 2008.
6. Abramov V. M. Asymptotic methods for queueing system and networks with application to telecommunications. Monash University, 2004.
7. Миллер А. Б. Динамическое управление доступом и скоростью обслуживания при активных пользователях.// Автоматика и телемеханика, 2010, № 9. С. 70-82.
8. Мясников Д. В., Семенихин К. В. Управление одноканальной системы массового обслуживания при наличии ограничений// Изв. РАН. Теория и системы управления, 2016, № 1, С. 66-85.
9. Миллер Д. М., Миллер Г. Б., Семенихин К. В. Регуляризация задачи оптимального управления марковским процессом с конечным числом состояний при наличии ограничений.// Автоматика и телемеханика, 2016, № 9, С. 96-123.
10. Alfonsi A. Affine Diffusions and Related Processes: Simulation, Theory and Applications. Springer, 2015.
11. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. М.: Наука, 1987.
12. Floyd S., Jacobson V. Random early detection gateways for congestion avoidance.// IEEE/ACM Transactions on Networking, 1993, Vol. 1, no. 4, Pp. 397-413.
13. Васенин В. А., Симонова Г. И. Математические модели управления трафиком в Интернет: новые подходы на основе схем TCP/AQM // Автоматика и телемеханика, 2005, № 8, С. 94-107.

Работа поступила 20.02.2019г.