На правах рукописи

Долгов Виталий Игоревич

ДИНАМИЧЕСКОЕ УПРАВЛЕНИЕ ИНТЕНСИВНОСТЯМИ ОБСЛУЖИВАНИЯ В СЕТЯХ МАССОВОГО ОБСЛУЖИВАНИЯ

01.01.09 – Дискретная математика и математическая кибернетика

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Саратов – 2010

Работа выполнена на кафедре системного анализа и автоматического управления Саратовского государственного университета им. Н. Г. Чернышевского

Научный руководитель:	доктор технических наук, профессор Митрофанов Юрий Иванович
Официальные оппоненты:	доктор физико-математических наук, профессор Розен Виктор Владимирович
	кандидат физико-математических наук, доцент Шульга Татьяна Эриковна
Ведущая организация:	Институт проблем точной механики и управления РАН

Защита состоится « 28 » октября 2010 г. в 17 часов 00 мин на заседании диссертационного совета ДМ 212.243.15 в Саратовском государственном университете им. Н. Г. Чернышевского по адресу: 410012, г. Саратов, ул. Астраханская, 83, механико-математический факультет.

С диссертацией можно ознакомиться в Научной библиотеке Саратовского государственного университета им. Н. Г. Чернышевского.

Автореферат разослан «____» сентября 2010 г.

Ученый секретарь
диссертационного совета
к.ф.-м.н., доцент В. В. Корнев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Эффективное использование сетей массового обслуживания в качестве математических моделей дискретных систем с сетевой структурой и стохастическим характером функционирования, примерами которых являются информационно-вычислительные сети, сети передачи данных, транспортные и гибкие производственные системы, обусловило продолжающееся более полувека интенсивное развитие теории сетей массового обслуживания, методов анализа, синтеза и оптимизации сетей массового обслуживания различных классов [1–9]. Как модели дискретных систем сети массового обслуживания используются для вычисления временных характеристик, коэффициентов использования устройств, надежности, производительности и других функциональных характеристик дискретных систем при достаточно общих предположениях об их структуре и процессах функционирования. Широкому практическому применению сетей массового обслуживания способствует простота и естественность, с которыми они отображают структуру моделируемых систем и процессы обработки в системах объектов различных типов. Большой вклад в развитие теории, методов анализа, оптимизации и синтеза сетей массового обслуживания внесли Г. П. Башарин, А. А. Боровков, П. П. Бочаров, В. М. Вишневский, В. А. Жожикашвили, В. А. Ивницкий, Ю. И. Митрофанов, В. В. Рыков. Среди зарубежных специалистов необходимо отметить значительный вклад в развитие этого научного направления таких ученых, как Дж. Джексон, Л. Клейнрок, Ф. Келли, К. Чэнди, Д. Тауслей, М. Райзер, Дж. Уолрэнд.

Отображение в модельных сетях массового обслуживания средств и методов управления дискретными системами приводит к построению сетей обслуживания с управлением, являющихся фактически подклассом сетей массового обслуживания. Сети обслуживания с управлением обеспечивают не только принципиальную возможность решения целого класса задач анализа и синтеза дискретных систем, но и возможность решения ряда задач, связанных с повышением эффективности управления дискретными системами.

Разработка и исследование методов управления интенсивностями обслуживания в сетях массового обслуживания и методов анализа сетей обслуживания с управлением интенсивностями обслуживания являются актуальными направлениями развития теории сетей массового обслуживания. Практическое значение этих направлений определяется широким использованием сетей массового обслуживания в качестве математических моделей дискретных систем с управлением, сетевой структурой и стохастическим характером функционирования. Интенсивности обслуживания требований системами обслуживания, входящими в состав сетей массового обслуживания, являются параметрами, в существенной степени определяющими качество функционирования сетей. Поэтому проблемам, связанным с исследованием влияния интенсивностей обслуживания на функционирование сетей обслуживания, определением оптимальных интенсивностей обслуживания, управлением интенсивностями обслуживания уделяется значительное внимание в современной теории сетей массового обслуживания.

В основу диссертации положены результаты научных исследований, выполненных при участии автора в Саратовском государственном университете по темам, включенным в план НИР СГУ: «Динамическое управление сетями массового обслуживания» (шифр «Темп», гос. рег. № 01200201953), «Анализ сетей массового обслуживания с динамическим управлением» (шифр «Тракт», гос. рег. № 01200602692), «Разработка и применение фундаментальных методов исследования задач математического анализа, дифференциальных уравнений, дискретной математики, теории упругости и газодинамики» (шифр «Интеграл», гос. рег. № 01200002986).

Цель диссертационной работы. Развитие теории сетей массового обслуживания с управлением и методов их анализа, разработка эффективных методов динамического управления интенсивностями обслуживания в сетях массового обслуживания, в том числе:

1. Разработка и исследование методов динамического управления интенсивностями обслуживания в замкнутых экспоненциальных сетях массового обслуживания с произвольной и типовой структурами.

2. Разработка и исследование методов анализа замкнутых экспоненциальных сетей массового обслуживания с произвольной и типовой структурами и динамическим управлением интенсивностями обслуживания для стационарного режима эволюции сетей.

3. Исследование зависимости стационарных характеристик сетей массового обслуживания от интенсивностей обслуживания.

Методы исследования. Использовались результаты теории вероятностей, теории марковских процессов, теории массового обслуживания, теории сетей массового обслуживания.

Основные результаты и научная новизна.

1. Разработаны методы динамического управления интенсивностями обслуживания в замкнутых экспоненциальных сетях массового обслуживания с одним классом требований и произвольной и типовой структурами.

2. Разработаны методы анализа замкнутых экспоненциальных сетей массового обслуживания с одним классом требований, с произвольной и типовой структурами и динамическим управлением интенсивностями обслуживания.

3. Проведено исследование эффективности методов динамического управления интенсивностями обслуживания в замкнутых экспоненциальных сетях массового обслуживания с произвольной и типовой структурами.

Полученные в диссертационной работе результаты являются новыми.

Теоретическая и практическая значимость. Работа носит теоретический характер. Научные результаты диссертационной работы представляют вклад в развитие теории сетей массового обслуживания с управлением интенсивностями обслуживания и методов их анализа.

Практическая значимость представленных в диссертационной работе результатов заключается в возможности применения рассмотренных методов динамического управления интенсивностями обслуживания в сетях массового обслуживания и методов анализа сетей массового обслуживания с управлением в математических моделях дискретных систем с сетевой структурой и стохастическим характером функционирования. Использование моделей этого вида позволит расширить круг задач анализа систем этого класса и повысить эффективность их решения.

Апробация работы. Результаты докладывались и обсуждались на научных семинарах кафедры системного анализа и автоматического управления Саратовского государственного университета, Международных научных конференциях «Компьютерные науки и информационные технологии» (1–4 июля 2007 года, 1–4 июля 2009 года, г. Саратов), Десятом Всероссийском симпозиуме по прикладной и промышленной математике (19–24 мая 2009 года, г. Санкт-Петербург), Ежегодных межвузовских научных конференциях «Компьютерные науки и информационные технологии» (27 апреля 2005 года, 19 мая 2006 года, г. Саратов), представлены и обсуждались на Шестом Всероссийском симпозиуме по прикладной и промышленной математике (1–7 октября 2005 года, г. Сочи–Дагомыс).

Публикации. По результатам диссертации опубликовано 9 работ.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав и заключения. Объем диссертации 109 страниц. Диссертация содержит 11 таблиц. Список литературы включает 82 наименования.

СОДЕРЖАНИЕ РАБОТЫ

Введение содержит общую характеристику работы.

В **первой главе** представлен обзор основных результатов по теории сетей массового обслуживания с управлением интенсивностями обслуживания, методам управления интенсивностями обслуживания в сетях массового обслуживания, методам построения оптимальных правил выбора интенсивностей обслуживания в сетях массового обслуживания, а также методам анализа сетей массового обслуживания с интенсивностями обслуживания, зависящими от состояния сети, и анализа сетей с управлением интенсивностями обслуживания.

Во **второй главе** предлагаются метод динамического управления интенсивностями обслуживания в сетях массового обслуживания с произвольной структурой и метод анализа сетей массового обслуживания этого типа.

Рассматривается замкнутая экспоненциальная сеть массового обслуживания N, образованная L системами массового обслуживания S_i, $i=1,\dots,L$, типа $M/M/1$ с интенсивностями обслуживания μ_i и содержащая Q требований одного класса. Переходы требований между системами обслуживания в процессе функционирования сети определяются неприводимой маршрутной матрицей $\Theta = (\theta_{ij})$, $i, j = 1,\dots,L$. Вектор $s^{(n)} = (s_i^{(n)})$, $i = 1,\dots,L$, где $s_i^{(n)}$ – число требований, находящихся в системе S_i, определяет состояние сети с номером n. Множество X состояний сети имеет мощность $c_X = |X|$. Обозначим через $I = \{1,\dots,L\}$ и $B = \{1,\dots,c_X\}$ соответственно множества номеров систем массового обслуживания и номеров состояний сети. Предполагается, что для каждого состояния $s^{(n)} \in X$ определено неотрицательное вещественное число $V^{(n)}$, называемое потенциалом этого состояния. Значение потенциала состояния отражает уровень значимости пребывания сети в этом состоянии с точки зрения обеспечения заданных характеристик качества ее функционирования. Нумерация состояний производится по убыванию потенциалов. Базовое состояние $s^{\circ} = (s_i^{\circ})$, $i = 1,\dots,L$, имеет номер 1 и наибольший потенциал. Выбор некоторого состояния сети в качестве базового обуславливается необходимостью достижения требуемых значений заданной стационарной характеристики сети, например, пропускной способности сети, общих задержек обслуживания или коэффициентов использования ресурсов сети массового обслуживания.

Множество X делится на подмножества Y и Z доминантных и ординарных состояний, $c_Y = |Y|$ и $c_Z = |Z|$. Множество номеров доминантных состояний сети обозначим через D. К доминантным относятся состояния, пребывание в которых обеспечивает значение основной характеристики качества функционирования сети более близкое к экстремальному значению, чем пребывание в ординарных состояниях. Потенциалы доминантных состояний превосходят по величине потенциалы ординарных состояний. При определении потенциалов состояний, способа нумерации состояний в множестве X и формировании множеств Y и Z используются векторы $d^{(n)} = (d_i^{(n)})$, $d_i^{(n)} = s_i^{(n)} - s_i^{\circ}$, и вектор $b = (b_i)$, $b_i > 0$, $i = 1,\dots,L$, граничных значений числа требований в системах обслуживания. Состояния $s^{(n)} \in X$, для которых справедливы неравенства $\left| d_i^{(n)} \right| \le b_i$, $i = 1,\dots,L$, относятся к множеству Y, остальные состояния – к множеству Z.

Пусть N^c – сеть массового обслуживания, структура, параметры и алгоритмы функционирования которой такие же, как у сети N, и которая отличается от N только тем, что в N^c реализовано динамическое управление интенсивностями обслуживания. Управление осуществляется посредством управляющих воздействий, формируемых в процессе функционирования сети N^c системой управления. Основной целью управления является достижение максимального значения стационарной вероятности $\pi(Y)$ пребывания сети в подмножестве Y доминантных состояний при ограничении $\overline{R}$ на интенсивность управления R, определяемую как число управляющих воздействий, формируемых в единицу времени.

Различаются два режима функционирования сети N^c – нормальный и коррективный. Периоды функционирования сети в этих режимах называются соответственно нормальным и коррективным тактами. Такт будем обозначать через $x^{(k)}$, где $k \in \{1, 2, \ldots\}$ – номер такта в общей последовательности тактов, а момент окончания такта $x^{(k)}$ – через $\tau^{(k)}$. Все такты имеют фиксированную длительность φ. Режимы функционирования отличаются используемыми в сети векторами интенсивностей обслуживания требований – в нормальном такте используется вектор $\mu = (\mu_i)$, $i = 1, \ldots, L$, в коррективном такте – коррективный вектор, зависящий от состояния сети в момент окончания предшествующего такта.

Вектор интенсивностей обслуживания, используемый в течение такта $x^{(k)}$, обозначим через $\mu^{(k)} = (\mu_i^{(k)})$, $i = 1, \ldots, L$, $k \in \{1, 2, \ldots\}$. Значения компонент вектора интенсивностей обслуживания $\mu^{(k+1)}$, используемого в течение такта $x^{(k+1)}$, определяются в зависимости от состояния сети $s^{(\cdot,k)}$ в момент $\tau^{(k)}$ и значения вектора b.

В момент $\tau^{(k)}$ выполняются следующие действия: 1) идентификация состояния $s^{(\cdot,k)}$; 2) формирование вектора $d^{(\cdot,k)} = (d_i^{(\cdot,k)})$, $i = 1, \ldots, L$, где $d_i^{(\cdot,k)} = s_i^{(\cdot,k)} - s_i^{\circ}$; 3) проверка выполнения неравенств $\left| d_i^{(\cdot,k)} \right| > b_i$. Если данные неравенства не выполняются для всех систем, то следующий такт является нормальным, и в течение такта $x^{(k+1)}$ используется вектор $\mu^{(k+1)} = \mu$. Если хотя бы для одной системы неравенство выполнилось, то следующий такт является коррективным, формируется вектор $\mu^{(k+1)} = \tilde{\mu}^{(k+1)}$, значения компонент которого направляются системам обслуживания и используются в течение такта $\tilde{x}^{(k+1)}$. После окончания такта $x^{(k+1)}$ в момент $\tau^{(k+1)}$ производится очередное выполнение действий 1) – 3) и т. д.

Предлагается следующий **метод формирования коррективного вектора интенсивностей обслуживания** $\tilde{\mu}^{(k+1)} = (\tilde{\mu}_i^{(k+1)})$, $i = 1, \ldots, L$. Обозначим через $\nu_{ij}^{(k+1)}$, $i, j \in I$, интенсивность потока требований из S_i в S_j в течение коррективного такта $\tilde{x}^{(k+1)}$, а через $\nu_i^{(k+1)}$ и $\upsilon_i^{(k+1)}$ – интенсивности потоков требований, выходящего из S_i и входящего в S_i в течение этого такта. Алгоритм формирования $\tilde{\mu}^{(k+1)}$ содержит вспомогательный шаг и основные шаги, число которых зависит от требуемой точности определения вектора $\tilde{\mu}^{(k+1)}$. При выполнении вспомогательного шага предполагается, что $\upsilon_i^{(k+1)} = \mu_i$, $i = 1, \ldots, L$, и определяются величины

$$\breve{\mu}_i^{(k+1)} = \begin{cases} (s_i^{(\cdot,k)} + \mu_i\varphi - s_i^{\circ})/\varphi & \text{при } s_i^{(\cdot,k)} + \mu_i\varphi > s_i^{\circ}, \\ 0 & \text{при } s_i^{(\cdot,k)} + \mu_i\varphi \le s_i^{\circ}. \end{cases}$$

На первом основном шаге полагается, что $\nu_i^{(k+1)} = \breve{\mu}_i^{(k+1)}$, $i = 1,\ldots,L$, и определяются

$$\nu_{ij}^{(k+1)} = \nu_i^{(k+1)}\theta_{ij},\ i,j = 1,\ldots,L, \quad \upsilon_j^{(k+1)} = \sum_{i=1}^{L}\nu_{ij}^{(k+1)},\ j = 1,\ldots,L,$$

$$\tilde{\mu}_i^{(k+1)} = \begin{cases} (s_i^{(\cdot,k)} + \upsilon_i^{(k+1)}\varphi - s_i^{\circ})/\varphi & \text{при } s_i^{(\cdot,k)} + \upsilon_i^{(k+1)}\varphi > s_i^{\circ}, \\ 0 & \text{при } s_i^{(\cdot,k)} + \upsilon_i^{(k+1)}\varphi \le s_i^{\circ}. \end{cases}$$

При выполнении следующих основных шагов производится подстановка вместо $\nu_i^{(k+1)}$ величин $\tilde{\mu}_i^{(k+1)}$, $i = 1,\ldots,L$, определенных на предыдущем основном шаге.

Теорема 2.9. *Оптимальная длительность такта, при которой достигается* $max\,\pi(Y)$ *при* $R \le \overline{R}$, $\varphi_0 = 1/\overline{R}$.

Замечание 2.2. *При* $\varphi \to \infty$ $\breve{\mu}_i^{(k+1)} \to \mu_i$ *и* $\tilde{\mu}_i^{(k+1)} \to \mu_i$, $i = 1,\ldots,L$.

Далее коррективный вектор интенсивностей обслуживания будем обозначать через $\tilde{\mu}^J$, $J \in \{1,\ldots,c_Z\}$, если коррективный такт начинается в состоянии $s^{(c_Y+J)} \in Z$.

Эволюция сети обслуживания N^c описывается случайным процессом Ξ с множеством состояний X, который является последовательностью фрагментов, соответствующих нормальным и коррективным тактам. Цепь Маркова, описывающую эволюцию сети N^c в течение нормальных тактов, обозначим через $\hat{C}$, а цепи Маркова, описывающие эволюцию сети N^c в течение коррективных тактов, – через $\tilde{C}^J$, $J \in \{1,...,c_Z\}$. Длительности реализаций цепей $\hat{C}$ и $\tilde{C}^J$ равны φ. Множеством начальных состояний цепи $\hat{C}$ является множество $\{s^{(1)},\ldots,s^{(c_Y)}\}$, а цепи $\tilde{C}^J$ – $\{s^{(c_Y+J)}\}$. Длительности пребывания цепей $\hat{C}$ и $\tilde{C}^J$ в состоянии $s^{(n)} \in X$ являются случайными величинами с экспоненциальными функциями распределения с параметрами $\hat{\alpha}_n$ и $\tilde{\alpha}_n^J$ соответственно. Для цепей $\hat{C}$ и $\tilde{C}^J$ соответственно обозначим $\hat{P} = (\hat{p}_{mn})$ и $\tilde{P}^J = (\tilde{p}_{mn}^J)$, $m,n = 1,...,c_X$, – матрицы вероятностей перехода; $\hat{P}^{(t)} = (\hat{p}_{mn}^{(t)})$ и $\tilde{P}^{(t),J} = (\tilde{p}_{mn}^{(t),J})$ – матрицы вероятностей перехода за время t.

Теорема 2.10. *Существует единственное стационарное распределение процесса* Ξ $\pi = (\pi_n)$, $n = 1,...,c_X$, *зависящее от* φ *и являющееся решением системы линейных уравнений*

$$\pi_n = \sum_{l \in D}\pi_l\,\hat{p}_{ln}^{(\varphi)} + \sum_{J=1}^{c_Z}\pi_{c_Y+J}\,\tilde{p}_{c_Y+J,n}^{(\varphi),J},\ n = 1,\ldots,c_X,$$

с условием нормировки $\sum_{n \in B}\pi_n = 1.$

Обозначим через $q_l(t)$ вероятность того, что процесс Ξ в момент $t \in (0, \varphi]$ находится в состоянии $s^{(l)} \in X$,

$$q_l(t) = \sum_{m \in D} \pi_m \hat{p}_{ml}^{(t)} + \sum_{J=1}^{c_Z} \pi_{c_Y+J} \tilde{p}_{c_Y+J,l}^{(t),J} .$$

Теорема 2.11. *Стационарные вероятности p_n перехода процесса Ξ в состояние $s^{(n)} \in X$, зависящие от длительности такта φ, определяются выражениями*

$$p_n = \frac{1}{\varphi} \int_0^{\varphi} \frac{\eta_n(t)}{\eta(t)} dt, \ n \in B,$$

где

$$\eta_n(t) = \sum_{l \in B} q_l(t) (\hat{\alpha}_l \hat{p}_{ln} \sum_{m \in D} \pi_m + \sum_{J=1}^{c_Z} \tilde{\alpha}_l^J \pi_{c_Y+J} \tilde{p}_{ln}^J),$$

$$\eta(t) = \sum_{l \in B} q_l(t) (\hat{\alpha}_l \sum_{m \in D} \pi_m + \sum_{J=1}^{c_Z} \tilde{\alpha}_l^J \pi_{c_Y+J}).$$

Обозначим: $\bar{s}_i$ – математическое ожидание числа требований в S_i; $\overline{\lambda}_i$ – интенсивность входящего в систему S_i потока требований; $\bar{u}_i$ – математическое ожидание длительности пребывания требований в S_i; ψ_i – коэффициент использования обслуживающего прибора в S_i.

Теорема 2.12. *Сеть N^c имеет следующие стационарные характеристики, $i = 1, \ldots, L$:*

$$\bar{s}_i = \sum_{k=1}^{Q} k P\{s_i = k\}, \text{ где } P\{s_i = k\} = \sum_{s^{(n)} \in X \,\&\, s_i^{(n)} = k} \pi_n ;$$

$$\overline{\lambda}_i = \overline{\mu}_i (1 - P\{s_i = 0\}), \text{ где } \overline{\mu}_i = \mu_i \sum_{m \in D} \pi_m + \sum_{J=1}^{c_Z} \tilde{\mu}_i^J \pi_{c_Y+J} ;$$

$$\bar{u}_i = \bar{s}_i / \overline{\lambda}_i ; \ \psi_i = \overline{\lambda}_i / \overline{\mu}_i = 1 - P\{s_i = 0\}.$$

Результаты численного и имитационного моделирования показывают, что применение в сетях обслуживания с произвольной структурой предложенного метода управления интенсивностями обслуживания обеспечивает возможность повышения качества функционирования сетей по сравнению с качеством функционирования сетей при отсутствии управления.

В **третьей главе** предлагаются метод динамического управления интенсивностями обслуживания в сетях массового обслуживания с типовой структурой и методы анализа сетей массового обслуживания этого типа.

Рассматривается замкнутая экспоненциальная сеть массового обслуживания N с Q требованиями одного класса. Отличие от сети, рассмотренной в главе 2, состоит в том, что в состав сети N помимо L систем обслуживания S_i, $i = 1, \ldots, L$, типа $M/M/1$ с интенсивностями обслуживания μ_i

входит система обслуживания S_0 типа $M/M/Q$ с интенсивностью μ_0 обслуживания требований одним прибором. Системы S_i, $i=1,\ldots,L$, называются базисными, а S_0 – терминальной. Переходы требований между системами в процессе функционирования сети N определяются неприводимой маршрутной матрицей $\Theta=(\theta_{ij})$, $i,j=0,1,\ldots,L$. Обозначим через $I=\{0,1,\ldots,L\}$ – множество номеров систем в сети, H – базисную подсеть, $J=\{1,\ldots,L\}$ – множество номеров систем в подсети H.

Обозначим: $\mu=(\mu_i)$, $i=1,\ldots,L$; $\bar{\zeta}(\mu)$ – математическое ожидание длительности реакции сети обслуживания для системы S_0; $c_i(\mu_i)$ – стоимость системы S_i при использовании в ней интенсивности обслуживания μ_i; $C(\mu)$ – суммарная стоимость сети N (не включая стоимости системы S_0), $C(\mu)=\sum_{i=1}^{L}c_i(\mu_i)$; λ_i – интенсивность входящего потока требований в систему S_i; ω_i – относительная интенсивность входящего потока требований в систему S_i, $\omega=(\omega_i)$, $i=0,1,\ldots,L$, – решение уравнения $\omega=\omega\Theta$ с условием $\sum_{i=0}^{L}\omega_i=1$.

Для сети N при различных ограничениях на интенсивности обслуживания и различных видах функций стоимости систем решаются следующие задачи оптимизации интенсивностей обслуживания.

А1. Минимизация функции стоимости $C(\mu)$ сети при ограничении $\bar{\zeta}(\mu)\le U$ $(U>0)$.

А2. Минимизация математического ожидания $\bar{\zeta}(\mu)$ длительности реакции сети при ограничении $C(\mu)\le \mathrm{B}$ $(\mathrm{B}>0)$.

Рассмотрим задачу А1 при $0<\mu_i<\infty$ и линейных функциях $c_i(\mu_i)=a_i\mu_i$, $a_i>0$, $i=1,\ldots,L$. Следующая теорема определяет решение $\hat{\mu}=(\hat{\mu}_i)$, $i=1,\ldots,L$, этой задачи оптимизации.

Теорема 3.1. *При заданном ограничении* $\bar{\zeta}(\mu)=U$ *функция* $C(\mu)=\sum_{i=1}^{L}a_i\mu_i$, *где* $a_i>0$, *принимает минимальное значение, если*

$$\hat{\mu}_i=\lambda_i+\frac{\sqrt{\lambda_i}}{\lambda_0 U\sqrt{a_i}}\sum_{j=1}^{L}\sqrt{a_j\lambda_j},\ i=1,\ldots,L,$$

где

$$\lambda_i=\frac{Q\omega_i}{\omega_0\left(U+1/\mu_0\right)},\ i=0,1,\ldots,L.$$

Пусть N^c – сеть массового обслуживания, структура, параметры и алгоритмы функционирования которой такие же, как у сети N, и которая отличается от N только тем, что в N^c реализовано динамическое управление интенсивностями обслуживания в базисных системах. Для параметров и характеристик сети N^c используются введенные для соответствующих параметров и характеристик сети N обозначения с индексом «c». В частности, число требований в подсети H сетей N и N^c обозначается соответственно через Q_H и Q_H^c. Целью управления интенсивностями обслуживания в N^c является уменьшение математического ожидания длительности реакции $\overline{\zeta}^c$ сети N^c для системы S_0 и вероятности ρ^c пребывания этой сети в состояниях, в которых Q_H^c больше математического ожидания $\overline{Q}_H$ числа требований в подсети H сети N.

Рассмотрим метод динамического управления интенсивностями обслуживания в сети N^c. Так же, как и в главе 2, предполагается, что в сети обслуживания реализована централизованная система управления. Различаются нормальные и коррективные такты функционирования сети. Обозначим через $\mu^{(k)} = (\mu_i^{(k)})$, $i = 1, \ldots, L$, – вектор интенсивностей обслуживания, используемых в S_i в течение такта $x^{(k)}$, а через $Q_H^{(k)}$– число требований в подсети H в момент $\tau^{(k)}$ окончания такта $x^{(k)}$.

В момент $\tau^{(k)}$ выполняются следующие действия: 1) определение $Q_H^{(k)}$; 2) проверка выполнения неравенства $Q_H^{(k)} \le \overline{Q}_H$. Если неравенство выполняется, то такт $x^{(k+1)}$ считается нормальным и в течение этого такта используется вектор $\mu^{(k+1)} = \mu$, в противном случае этот такт – коррективный, формируется зависящий от $Q_H^{(k)}$ вектор $\mu^{(k+1)} = \tilde{\mu}^{(k+1)}$, значения компонент которого направляются базисным системам и используются в течение $\tilde{x}^{(k+1)}$. После окончания такта $x^{(k+1)}$ в момент $\tau^{(k+1)}$ производится очередное выполнение действий 1) и 2) и т. д.

Предлагается следующий **метод формирования коррективного вектора** $\tilde{\mu}^{(k+1)} = (\tilde{\mu}_i^{(k+1)})$, $i = 1, \ldots, L$. Значения $\tilde{\mu}_i^{(k+1)}$ находятся с использованием теоремы 3.1, как решение соответствующей задачи оптимизации,

$$\tilde{\mu}_i^{(k+1)} = \tilde{\lambda}_i^{(k+1)} + \frac{\sqrt{\tilde{\lambda}_i^{(k+1)}}}{\tilde{\lambda}_{H0}^{(k+1)} U \sqrt{a_i}} \sum_{j=1}^{L} \sqrt{a_j \tilde{\lambda}_j^{(k+1)}}\,,\ i \in J\,,$$

где $\tilde{\lambda}_{H0}^{(k+1)}$ – интенсивность потока требований из H в S_0,

$$\tilde{\lambda}_{H0}^{(k+1)} = \mu_0 (\bar{s}_0 - s_0^{(k)} e^{-\mu_0 \varphi}) / (1 - e^{-\mu_0 \varphi})\,,$$

а $\tilde{\lambda}_i^{(k+1)}$ – интенсивность потока требований в S_i,

$$\tilde{\lambda}_i^{(k+1)} = \tilde{\lambda}_{H0}^{(k+1)} \omega_i \big/ \omega_0\,.$$

Здесь $\bar{s}_0$ – математическое ожидание числа требований в S_0 сети N, $s_0^{(k)}$ – число требований в S_0 сети N^c в момент $\tau^{(k)}$.

Округлим с избытком $\overline{Q}_H$ до целого числа q. Для каждого из $\mathcal{D} = Q - q + 1$ значений $Q_H^{(k)} > \overline{Q}_H$ формируется используемый в течение коррективного такта вектор $\mu^{(r)}$, $r \in \{1,\ldots,\mathcal{D}\}$. Определим разбиение множества X состояний сети N^c на подмножества $X^{(0)}$, $X^{(1)}$, …, $X^{(\mathcal{D})}$ такие, что $X^{(0)} = \{s^{(n)} \in X \mid 0 \le Q_H^c \le q-1\}$, $X^{(j)} = \{s^{(n)} \in X \mid Q_H^c = j+q-1\}$, $j = 1,\ldots,\mathcal{D}$. Множество $X^{(j)}$, $j \in \{0,1,\ldots,\mathcal{D}\}$, назовем агрегированным состоянием с номером j сети N^c. Обозначим через $B^{(j)}$ множество номеров состояний сети N^c из множества $X^{(j)}$.

Предлагаются два метода анализа сети N^c с типовой структурой и динамическим управлением интенсивностями обслуживания. В основе первого метода анализа лежит предположение о достижении сетью к моменту окончания такта стационарного режима.

Пусть сети N и $N^{(r)}$, $r \in \{1,\ldots,\mathcal{D}\}$, с векторами интенсивностей обслуживания соответственно μ и $\mu^{(r)}$ отличаются от N^c только отсутствием в них управления. Стационарные вероятности $P(s)$ и $P^{(r)}(s)$, $s \in X$, состояний сетей N и $N^{(r)}$ определяются по формулам:

$$P(s) = \frac{1}{G s_0!}\prod_{i=0}^{L}\left(\frac{\omega_i}{\mu_i}\right)^{s_i}, \quad G = \sum_{s \in X}\frac{1}{s_0!}\prod_{i=0}^{L}\left(\frac{\omega_i}{\mu_i}\right)^{s_i};$$

$$P^{(r)}(s) = \frac{1}{G^{(r)} s_0!}\prod_{i=0}^{L}\left(\frac{\omega_i}{\mu_i^{(r)}}\right)^{s_i}, \quad G^{(r)} = \sum_{s \in X}\frac{1}{s_0!}\prod_{i=0}^{L}\left(\frac{\omega_i}{\mu_i^{(r)}}\right)^{s_i}.$$

Стационарные распределения $\gamma = (\gamma_j)$ и $\gamma^{(r)} = (\gamma_j^{(r)})$, $j = 0,1,\ldots,\mathcal{D}$, вероятностей агрегированных состояний сетей N и $N^{(r)}$ определяются выражениями:

$$\gamma_j = \sum_{s \in X^{(j)}} P(s), \quad \gamma_j^{(r)} = \sum_{s \in X^{(j)}} P^{(r)}(s).$$

Теорема 3.5. *Стационарное распределение* $\chi = (\chi_j)$, $j = 0,1,\ldots,\mathcal{D}$, *вероятностей агрегированных состояний сети* N^c *является решением уравнения*

$$\chi\Gamma = \chi,$$

где

$$\Gamma = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{\mathcal{D}} \\ \gamma_0^{(1)} & \gamma_1^{(1)} & \cdots & \gamma_{\mathcal{D}}^{(1)} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_0^{(\mathcal{D})} & \gamma_1^{(\mathcal{D})} & \cdots & \gamma_{\mathcal{D}}^{(\mathcal{D})} \end{pmatrix},$$

с условием нормировки $\sum_{j=0}^{\mathcal{D}} \chi_j = 1$.

Следствие 3.1. *Если известны стационарные характеристики* $\overline{g}_i$ *и* $\overline{g}_i^{(r)}$, $r=1,\dots,\mathcal{D}$, *сетей* N *и* $N^{(r)}$, *то соответствующая стационарная характеристика сети* N^c

$$\overline{g}_i^c = \overline{g}_i \chi_0 + \sum_{r=1}^{\mathcal{D}} \overline{g}_i^{(r)} \chi_r .$$

Поскольку в основе данного метода анализа сети N^c лежит предположение о достижении сетью к моменту окончания такта стационарного режима, то при использовании этого метода необходимо задать величину ε ошибки приближения эволюции сети N^c к стационарному режиму.

Теорема 3.6. *При* $\varphi \geq t_e = \ln(Q/\varepsilon)/\mu_0$ *режим эволюции сети* N^c *является стационарным с известной ошибкой* ε.

Таким образом, предлагаемый метод анализа применим при таких длительностях тактов, в моменты окончания которых сеть функционирует в режиме, близком к стационарному.

В основе второго предлагаемого метода анализа лежит описание эволюции сети N^c случайным процессом Ξ с непрерывным временем и конечным множеством состояний B, который представляет собой последовательность фрагментов, соответствующих нормальным и коррективным тактам функционирования сети (аналогично описанию в главе 2). Фрагментами реализаций процесса Ξ являются реализации цепей Маркова $\hat{C}$ и $\tilde{C}^{(r)}$, $r \in \{1,\dots,\mathcal{D}\}$, с множеством состояний B и непрерывным временем. Эволюция сети N^c в течение нормальных тактов описывается цепью $\hat{C}$ с множеством начальных состояний $B^{(0)}$, а в течение коррективных тактов – цепями $\tilde{C}^{(r)}$ с множествами начальных состояний $B^{(r)}$ соответственно. Длительности реализаций цепей $\hat{C}$ и $\tilde{C}^{(r)}$ равны φ. Обозначим через $\hat{P}^{(t)} = (\hat{p}_{mn}^{(t)})$, $\tilde{P}^{(t),(r)} = (\tilde{p}_{mn}^{(t),(r)})$, $m, n = 1,\dots,c_X$, матрицы вероятностей перехода за время t цепей Маркова $\hat{C}$ и $\tilde{C}^{(r)}$ соответственно.

Теорема 3.7. *При заданном значении* φ *стационарное распределение* $\pi = (\pi_n)$, $n = 1,\dots,c_X$, *сети* N^c *существует, является единственным и удовлетворяет уравнению*

$$\pi = \pi P,$$

где $P = (p_{mn})$, $m, n = 1,\dots,c_X$,

$$p_{mn} = \begin{cases} \hat{p}_{mn}^{(\varphi)}, & m \in B^{(0)}, \\ \tilde{p}_{mn}^{(\varphi),(r)}, & m \in B^{(r)}, r \in \{1,\dots,\mathcal{D}\}. \end{cases}$$

Очевидно, $\rho^c = \sum_{r=1}^{D} \sum_{s^{(n)} \in X^{(r)}} \pi_n$.

Стационарные характеристики сети N^c вычисляются по формулам, приведенным в теореме 2.12, с использованием математических ожиданий интенсивностей обслуживания в базисных системах

$$\overline{\mu}_i^c = \mu_i \sum_{s^{(n)} \in X^{(0)}} \pi_n + \sum_{r=1}^{\mathcal{D}} \mu_i^{(r)} \sum_{s^{(n)} \in X^{(r)}} \pi_n \,,\; i = 1, \ldots, L \,.$$

Математическое ожидание длительности реакции сети N^c для S_0

$$\overline{\zeta}^c = \frac{1}{\mu_0 \overline{s}_0^c} \sum_{i=1}^{L} \overline{s}_i^c \,.$$

Как показывают результаты численного и имитационного моделирования, предложенные методы анализа обеспечивают возможность определения с удовлетворительной точностью стационарных характеристик замкнутых экспоненциальных сетей массового обслуживания с типовой структурой и динамическим управлением интенсивностями обслуживания при различных длительностях тактов. Длительность такта в сетях этого типа, являясь одним из основных параметров метода управления, в существенной степени определяет качество функционирования сетей.

В **четвертой главе** рассматриваются принципы организации, структура и основные алгоритмы функционирования имитационных моделей, разработанных для анализа сетей массового обслуживания с различной структурой и динамическим управлением интенсивностями обслуживания. Описываются методы моделирования случайных величин и анализа результатов имитационного моделирования.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработаны методы динамического управления интенсивностями обслуживания в замкнутых экспоненциальных сетях массового обслуживания с произвольной и типовой структурами.

2. Разработаны методы анализа замкнутых экспоненциальных сетей массового обслуживания с произвольной и типовой структурами и динамическим управлением интенсивностями обслуживания для стационарного режима эволюции сетей.

3. Методами аналитического и имитационного моделирования проведено исследование эффективности методов динамического управления интенсивностями обслуживания в сетях массового обслуживания с произвольной и типовой структурами.

СПИСОК ЛИТЕРАТУРЫ

1. Башарин, Г.П. Анализ очередей в вычислительных сетях. Теория и методы расчета / Г.П. Башарин, П.П Бочаров., Я.А. Коган. – М.: Наука. ГРФМЛ, 1989. – 336 с.

2. Боровков, А.А. Асимптотические методы в теории массового обслуживания / А.А. Боровков. – М.: Наука, 1980. – 384 с.

3. Жожикашвили, В.А. Сети массового обслуживания. Теория и применение в сетях ЭВМ / В.А. Жожикашвили, В.М. Вишневский. – М.: Радио и связь, 1988. – 192 с.

4. Клейнрок, Л. Теория массового обслуживания / Л. Клейнрок. – М.: Машиностроение, 1979. – 432 с.

5. Митрофанов, Ю.И. Анализ сетей массового обслуживания / Ю.И. Митрофанов. – Саратов: Научная книга, 2005. – 177 с.

6. Уолренд, Дж. Введение в теорию сетей массового обслуживания / Дж. Уолренд. – М.: Мир, 1993. – 336 с.

7. Jackson, J.R. Networks of waiting lines / J.R. Jackson // Oper. Res. – 1957. – V. 5, № 4. – P. 518-521.

8. Kelly, F.P. Reversibility and stochastic networks / F.P. Kelly. – London, Wiley, 1979. – 230 p.

9. Reiser, M. Mean-value analysis of closed multichain queueing networks / M. Reiser, S.S. Lavenberg // J. ACM. – 1980, V. 27, №. 2. – P. 313-322.

СПИСОК РАБОТ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ

10. Долгов, В.И. Метод анализа сетей массового обслуживания с динамическим управлением интенсивностями обслуживания / В.И. Долгов, Ю.И. Митрофанов, Е.С. Рогачко // Известия Сарат. ун-та. Нов. серия. Серия Математика. Механика. Информатика. – 2009. – Т. 9, вып. 3. – С. 22-27.

В. И. Долгову принадлежит метод анализа замкнутых экспоненциальных сетей массового обслуживания с динамическим управлением интенсивностями обслуживания.

11. Долгов, В.И. Модели и методы анализа сетей массового обслуживания с динамическим управлением интенсивностями обслуживания / В.И. Долгов, Ю.И. Митрофанов, Е.С. Рогачко // Обозрение прикладной и промышленной математики: Тез. докл. X Всерос. симпоз. по прикл. и пром. матем., Санкт-Петербург, 19–24 мая 2009 г. – М.: Науч. изд-во «ОПиПМ», 2009. – Т. 16, вып. 2. – С. 323-324.

В. И. Долгову принадлежат методы анализа сетей массового обслуживания с динамическим управлением интенсивностями обслуживания.

12. Долгов, В.И. Исследование сетей массового обслуживания с управлением интенсивностями обслуживания / В.И. Долгов // Обозрение прикладной и промышленной математики: Тез. докл. VI Всерос. симпоз. по прикл. и пром. матем., Сочи–Дагомыс, 1–7 окт. 2005 г. – М.: Науч. изд-во «ОПиПМ», 2005. – Т. 12, вып. 4. – С. 949-950.

13. Долгов, В.И. Динамическое управление интенсивностями обслуживания в сетях массового обслуживания / Ю.И. Митрофанов, В.И. Долгов // Автоматика и вычислительная техника. – 2008. – № 6. – С. 44-56.

В. И. Долгову принадлежат метод динамического управления интенсивностями обслуживания в сетях массового обслуживания, метод анализа сетей массового обслуживания с управлением интенсивностями обслуживания и результаты исследования методами численного и имитационного моделирования сетей с управлением интенсивностями обслуживания.

14. Долгов, В.И. Сети массового обслуживания с управлением интенсивностями обслуживания: синтез, метод управления, исследование / Ю.И. Митрофанов, В.И. Долгов; Сарат. гос. ун-т. – Саратов, 2005. – 26 с. – Деп. в ВИНИТИ 13.05.05, № 688-В2005.

В. И. Долгову принадлежат метод динамического управления интенсивностями обслуживания в замкнутых экспоненциальных сетях массового обслуживания, метод формирования вектора минимальных интенсивностей обслуживания, структура и основные алгоритмы функционирования имитационной модели, разработанной для исследования эффективности метода управления.

15. Долгов, В.И. Исследование зависимости характеристик сетей массового обслуживания с управлением интенсивностями обслуживания от топологии сетей / В.И. Долгов; Сарат. гос. ун-т. – Саратов, 2005. – 23 с. – Деп. в ВИНИТИ 25.05.05, № 744-В2005.

16. Долгов, В.И. Исследование зависимости качества функционирования сети массового обслуживания с управлением интенсивностями обслуживания от реакции системы управления сетью / В.И. Долгов // Теоретические проблемы информатики и ее приложений: Сб. науч. тр. – Саратов: Изд-во Сарат. ун-та, 2006. – Вып. 7. – С. 47-51.

17. Долгов, В.И. Имитационная модель сети массового обслуживания с динамическим управлением интенсивностями обслуживания / В.И. Долгов // Компьютерные науки и информационные технологии: Матер. Междунар. науч. конф., Саратов, 1–4 июля 2009 г. – Саратов: Изд-во Сарат. ун-та, 2009. – С. 80-83.

18. Долгов, В.И. Моделирование сети массового обслуживания с управлением интенсивностями обслуживания / Ю.И. Митрофанов, В.И. Долгов, Е.П. Станкевич // Компьютерные науки и информационные технологии: Тез. докл. Междунар. науч. конф., посвящ. памяти проф. А.М. Богомолова, Саратов, 1–4 июля 2007 г. – Саратов: Изд-во Сарат. ун-та, 2007. – С. 83-85.

В. И. Долгову принадлежат метод динамического управления интенсивностями обслуживания в сетях массового обслуживания и метод анализа сетей массового обслуживания с управлением интенсивностями обслуживания.

Работы [10–12] опубликованы в журналах, включенных в перечень ВАК ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени кандидата наук.

Подписано в печать 17.09.2010 Формат 60x84 1/16. Бумага офсетная.
Гарнитура Times. Печать RISO. Объем 1,0 печ.л. Тираж 100 экз.

Отпечатано с готового оригинал-макета
Центр полиграфических и копировальных услуг
Предприниматель Серман Ю. Б. Свидетельство № 3117
410000, г. Саратов, ул. Московская, 152, офис 19