Некоммерческое акционерное общество
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ»

Кафедра «Телекоммуникационные системы»
Специальность 6М071900 «Радиотехника, электроника и телекоммуникации»

ДОПУЩЕН К ЗАЩИТЕ
Зав. кафедрой
к.т.н., Шагиахметов Д.Р.
(ученая степень, звание, ФИО) (подпись)
«____»_____2014 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ
пояснительная записка

на тему: Модели и методы мультисервисного контакт-центра

Магистрант Таскинбаев Н.К. _____ группа МТСп-12-2
(Ф.И.О.) (подпись)
Руководитель к.ф-м.н., доцент _____ Жунусов К.Х.
(ученая степень, звание) (подпись) (Ф.И.О.)

Рецензент к.т.н., профессор _____ Хасенова Г.И.
(ученая степень, звание) (подпись) (Ф.И.О.)

Консультант по ВТ к.х.н., ст.преп. _____ Данько Е.Т.
(ученая степень, звание) (подпись) (Ф.И.О.)

Нормоконтроль к.х.н., ст.преп. _____ Кудинова В. С.
(ученая степень, звание) (подпись) (Ф.И.О.)

Алматы, 2014

Некоммеческое акционерное общество
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ»

Факультет__«Радиотехники, электроники и связи»_____
Специальность_6М071900 «Радиотехники, электроники и телекоммуникации»
Кафедра__«Телекоммуникационных систем»_____

ЗАДАНИЕ
на выполнение магистерской диссертации

Магистранту__Таскинбаеву Н.К._____
(фамилия, имя, отчество)

Тема диссертации __« Модели и методы мультисервисного контакт – центра»_____
утверждена Ученым советом университета №142_от «_31»_октября 2013 г.___
Срок сдачи законченной диссертации «25»_декабря 2013г_____

Цель исследования___ Целью диссертационной работы является исследование моделей и методов расчета вероятностно-временных характеристик (ВВХ) мультисервисных контакт-центров, обеспечивающих приоритетную дисциплину обслуживания заявок и режим прямого и отложенного обслуживания.

_____

Перечень подлежащих разработке в магистерской диссертации вопросов или краткое содержание магистерской диссертации:
_1. Основные направления развития абонентского доступа_____
_2. Организация сетей доступа_____
_3. Оптические сети доступа_____
_4. **Технологические характеристики оптических сетей доступа**
_____ _5. Расчет основных параметров качества абонентской линии и результаты__
расчетов_____

Перечень графического материала (с точным указанием обязательных чертежей)_____ __
1. Обобщенная структура Call-центра_____
2. Модель центра обслуживания вызовов в виде СМО
3. Структурная схема оптической сети абонентского доступа по технологии PON_____
4. Схема построение мультисервисного Контакт - центра
5. Зависимость выходной мощности от входной мощности при λ=1550нм, L=100км, υ=2500Мбит/с _____

Рекомендуемая основная литература

1. Банкет В.Л., Бондаренко О.В. Современные телекоммуникации. Технологии и экономика. - М.: ЭКО_ТРЕНДЗ, 2001.

2. Фриман Р. Волоконно-оптические системы связи. – М.: ТЕХНОСФЕРА, 2003.

Г Р А Ф И К

подготовки магистерской диссертации

Наименование разделов, перечень разрабатываемых вопросов	Сроки представления научному руководителю	Примечание
1 Информационный обзор абонентских линий	05.10.2012	
2 Основные виды технологий абонентских линий	02.02.2013	
3 Анализ основных параметров сети PON	10.03.2013	
4 Оценка работы абонентских линий при различных технологиях и стандартах	05.09.2013	
5 Расчет различных параметров качества PON	18.10.2013	
6 Анализ полученных экспериментальных и расчетных данных	10.12.2013	

Дата выдачи задания_____

Заведующий кафедрой_____(__Коньшин С.В.___)

(подпись) (Ф.И.О.)

Руководитель диссертации_____(___Жунусов К.Х.____)

(подпись) (Ф.И.О.)

Задание принял к исполнению

магистрант _____(___Таскинбаев Н.К.__)

(подпись) (Ф.И.О.)

Содержание

Введение

В настоящее время внимание мирового телекоммуникационного сообщества сосредоточено на концепции сетей, которые обеспечивают предоставление любых услуг электросвязи на основе единой сетевой инфраструктуры, таких как сети следующего поколения (NGN). Усиление конкуренции в отрасли, а также повышение требований пользователей телекоммуникационных сетей привели к появлению качественно новых методов и средств предоставления услуг, основывающихся на конвергенции сетей связи и услуг. Одним из перспективных направлений развития информационных услуг является организация центров обслуживания вызовов (ЦОВ). Вместе с переходом от телефонных сетей общего пользования (ТфОП) к сетям следующего поколения (NGN) можно наблюдать эволюцию традиционных центров обслуживания вызовов (ЦОВ) к мультисервисным центрам обслуживания вызовов (МЦОВ) или мультисервисным контакт-центрам (МКЦ), обладающих несравненно большим набором услуг и возможностями. Задачей МКЦ, является предоставление пользователю любого удобного для него средства получения информационных услуг, будь то речевой или видео вызов, запрос по электронной почте или текстовый диалог, запрос из социальных сетей или прием заявок, допускающих отложенную обработку. Разнообразие типов обслуживаемых запросов приводит к существенным изменениям функциональной структуры рассматриваемого МКЦ, по сравнению с системами прошлого поколения. Эти принципиально новые подходы к предоставлению современных инфокоммуникационных услуг, ориентированных на сети связи следующего поколения (NGN), делают актуальными исследования моделей и методов проектирования таких центров.

Целью диссертационной работы является исследование моделей и методов расчета вероятностно-временных характеристик (ВВХ) мультисервисных контакт-центров, обеспечивающих приоритетную дисциплину обслуживания заявок и режим прямого и отложенного обслуживания.

Поставленная цель определила необходимость решения следующих задач:

а) разработка формализованного описания исследуемого объекта - мультисервисного контакт-центра (МКЦ);

б) исследование специфики процессов обслуживания запросов в МКЦ. Разработка методов расчетов ВВХ контакт-центра при обслуживании разнотипных потоков вызовов по приоритетной дисциплине;

в) исследование мультисервисного контакт-центра с отложенным обслуживанием запросов и разработка методов оценки её основных ВВХ;

г) разработка имитационной модели, позволяющей проводить оценку ВВХ мультисервисного контакт-центра с отложенным обслуживанием запросов;

д) разработка обобщенной методики проектирования мультисервисного контакт-центра.

Определению основных параметров, влияющих на качество предоставления информационных услуг, посвящен ряд научных работ [7, 8,], однако они ограничиваются, в основном, традиционными центрами обслуживания телефонных вызовов.

Следует отметить ряд работ [13,15], где исследуются подходы к получению характеристик более сложных систем, которые можно рассматривать как прообраз современных контакт-центров. Но эти работы не учитывают возможностью обслуживания изучаемыми системами потоков нагрузки, как напрямую, так и с отложенным, причем со всех видов обращений включая телефонный вызов. В связи с этим необходимо упомянуть работы [9,15], посвященные экспериментальному изучению характеристик нагрузки, которая может поступать на МКЦ с отложенным обслуживанием определяющие качество предоставления информационных услуг рассматриваемыми системами. Это позволяет эффективно решить проблему проектирования МКЦ, управления его работой в процессе эксплуатации и добиться положительного экономического эффекта. Результаты работы могут быть использованы научно-исследовательскими, производственными и эксплуатационными организациями при разработке, внедрении новых и усовершенствовании существующих центров информационных услуг.

1 Эволюция технологий и математических моделей контакт-центров

1.1 Эволюция центров предоставления информационных услуг от простейших систем распределения вызовов до IP-контакт- центров

Первые центры обслуживания вызовов (ЦОВ) строились на базе телефонных станций, интегрированных с системой автоматического распределения вызовов СРВ [1]; по сравнению с телефонными станциями эти центры были наделены более широкими функциональными возможностями обработки вызовов. Изначально ЦОВ предназначались для эффективного обслуживания большого количества однотипных телефонных вызовов при ограниченности ресурсов. Функциональность первых таких центров ограничивалась справочно-информационными службами сети общего пользования. Далее развитие ЦОВ шло по пути совершенствования системы маршрутизации вызовов (придания ей интеллектуальных черт) и системы голосовых меню IVR (Interactive Voice Response). Интеллект подобных систем ограничивался выдачей статистических отчётов об общей производительности ЦОВ, например, о числе вызовов на одного оператора в час.

В ранних версиях системы СРВ дисциплина выбора вызовов из очереди предусматривала маршрутизацию вызова, стоящего в очереди первым, к незанятому оператору, который был обнаружен первым при циклическом поиске. Такая дисциплина выбора вызовов работает хорошо, если поступающий трафик равномерен, а все операторы имеют одинаковую квалификацию; в противном случае её применение ведёт к перегрузке наиболее квалифицированного персонала. Если поступающий трафик неравномерен, а квалификация операторов различна, то целесообразно вызов, стоящий в очереди первым, маршрутизировать к терминалу того оператора, который простаивал дольше других; подобная стратегия позволяет распределить нагрузку между операторами более равномерно.

Традиционно вызовы, установленные в очередь, обрабатывались в соответствии с дисциплиной обслуживания FIFO - "первым поступил - первым обслужен". Однако, разнообразие задач, стоящих перед системой СРВ, приводит к модификациям дисциплины организации очередей с возможностью производить выбор вызовов из очереди не только в порядке их поступления.

В числе достоинств традиционных центров обработки вызовов, реализованных на базе телефонных станций, специалисты отмечают высокую надежность, проверенную на протяжении многих лет, и возможность использования уже имеющегося оборудования. Однако первоначально решения имели фиксированную функциональность и ограниченные средства интеграции с другими информационно-коммуникационными системами. Математические модели таких систем рассматривались в работах [5, 6, 7] и

др. Их положения и результаты, существенные для данной диссертационной работы, будут рассмотрены в следующих параграфах (1.3).

Основной задачей ЦОВ является обслуживание потока вызовов высокой интенсивности с минимальными потерями, для чего требуются гибкие алгоритмы распределения вызовов и процедуры их обслуживания. На сегодняшний день подавляющее большинство call-центров представляют собой систему автоматического распределения вызовов на базе автоматических телефонных станций или IP-коммутаторов.

Следующая ступень эволюции операторских центров - Call-центры. Все, что сказано о функциональных возможностях систем СРВ, относится и к call- центрам, однако отметим следующее. Система СРВ - это коммутационная система со специальными функциями, а call-центр - это учреждение, включающее систему СРВ, оснащенную оборудованием и специализированными программными средствами, и укомплектованный штат технического и управленческого персонала (рисунок 1.1). На смену работавшим в ЦОВ неквалифицированным операторам пришли обученные специалисты.

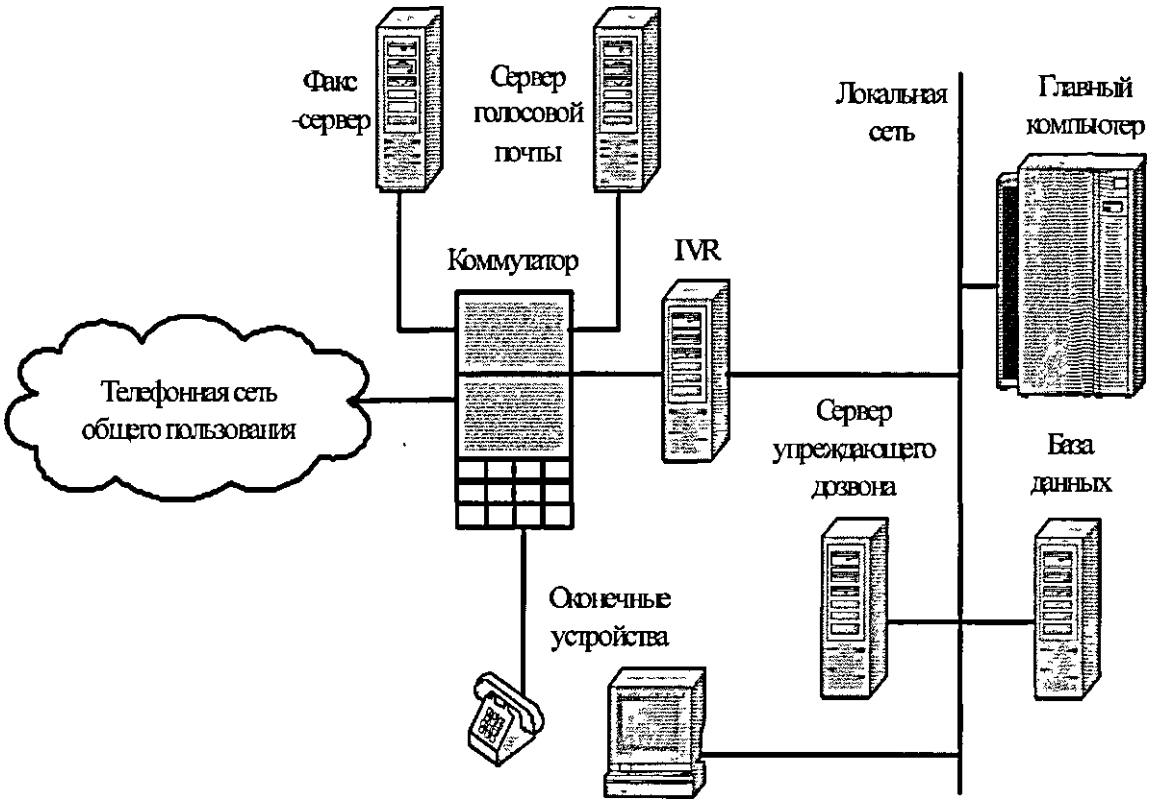


Рисунок 1.1 - Обобщенная структура Call-центра

Фактически одновременно с появлением первых компьютеров, возникла необходимость расширения функций ЦОВ, и производители телефонных станций начали увязывать эти решения с компьютерными приложениями (Computer Telephony Integration, CTI). Центры обслуживания вызовов на базе

коммуникационных платформ компьютерной телефонии предоставляют собой систему разнообразные готовые приложения и средства разработки собственных приложений, поддерживающих открытые телефонные протоколы и стандарты. Математические модели для таких архитектур будут рассматриваться в следующем параграфе.

Когда помимо телефонного звонка появилось множество других каналов для обращения в подобный центр, переход на унифицированную платформу обслуживания отразился в появлении нового названия - контакт-центр (КЦ). Современный контакт-центр - это интегрированные системы, взаимодействующие телекоммуникационными средами. Они поэтапно маршрутизируют входящие вызовы к наиболее подходящему по квалификации оператору и применяют технологии электронной коммерции, Web-запросы, обработку электронной почты, push-технологии, синхронную Web-навигацию, чаты, IP-телефонию.

На сегодняшний день большинство специалистов склонны выделять три основных типа мультимедийных запросов в адрес контакт-центра: голосовой (речевой), электронная почта, текстовый чат. Миграция функциональности ЦОВ от систем распределения вызовов к мультимедийным контакт-центрам нашла свое отражение во многих современных системах.

Сети нового поколения NGN (Next Generation Network) кардинальным образом меняют возможности контакт-центра, превращая его в точку входа в единое информационное пространство компании. Напомним, что сети NGN обеспечивают передачу голоса, данных и видеоизображения по одному каналу. В последнее время многие эксперты полагают, что мировой рынок контакт- центров достиг состояния зрелости - на нем представлены разнообразные решения различных производителей и выполнено множество проектов. Все более востребованными становятся IP-контакт-центры на базе пакетной коммутации, реализованные в программных продуктах технологии интеллектуальной маршрутизации. IP-контакт-центры (IPCC), представляют собой программные решения с органичным сочетанием традиционных и мультимедийных возможностей (наряду с телефонным вызовом они способны принимать запросы, отправляемые посредством электронной почты, HTTP и т. д.).

Интеграция таких систем с информационными системами предприятия достаточно проста благодаря изначальной ориентированности решений на базе IP на открытые стандарты и протоколы. Они могут быть установлены в любом месте и не зависят от операторов, агентов или от поставщиков услуг связи. Ни одна из сторон не привязана к телефонным окончаниям, и физически пользователи могут находиться везде, где есть канал IP. Более того, подключение не зависит от канала связи, это может быть телефонный канал, синхронный, Frame Relay, xDSL, ATM и т. д. лишь бы по нему передавались пакеты IP.

Собственно телефонный трафик можно получать посредством VoIP от любого оператора, а не только по телефонным каналам, при помощи, которых ЦОВ подключаются к традиционному оператору.

Управление работой IP-контакт-центра, изменение его функциональности и добавление новой выполняются через привычные и понятные интерфейсы, экранные формы и визуальные средства. Простота организации удаленных рабочих мест и масштабирование системы не зависят от наличия телефонных линий и портов в телефонной станции. Все чаще в них задействуются программные приложения, например, система IVR. Все отчетливее проявляется тенденция увеличения числа контакт-центров, имеющих сетевую архитектуру: в общем количестве постоянно множащихся операторских центров обслуживания вызовов их доля растет опережающими темпами. Сетевой контакт-центр - это не обязательно вынос крупных узлов, он может быть организован путем создания удаленных индивидуальных рабочих мест операторов.

1.2 Математические модели телефонных центров обслуживания вызовов

Исследованиям операторским центров и, в частности, ACD (автоматический распределитель вызовов) посвящен ряд публикаций. Часть разработанных научных методов исследований представляется полезной для исследования 1Р-контакт-центров

В зависимости от характеристик, оборудование центров предоставления информационных услуг может быть представлено в виде различных моделей систем массового обслуживания (СМО)[5], в том числе:

– модели СМО M/M/n/, M/M/n/К;

– модели СМО вида M1+.. .+Mc/M/n/К;

– модели СМО M/G/п или G/G/n/ с распределениями времен обслуживания заявок и их поступления, отличными от показательного и учитывающие свойства самоподобия процессов (Логнормальное, Парето и др.);

– модели СМО с различными дисциплинами приоритетов поступающих на обслуживание заявок.

Кратко остановимся на основных методах исследования центров обслуживания вызовов в соответствии с их эволюцией. На рисунке 1.2 приведен простая функциональная модель Call-центра.

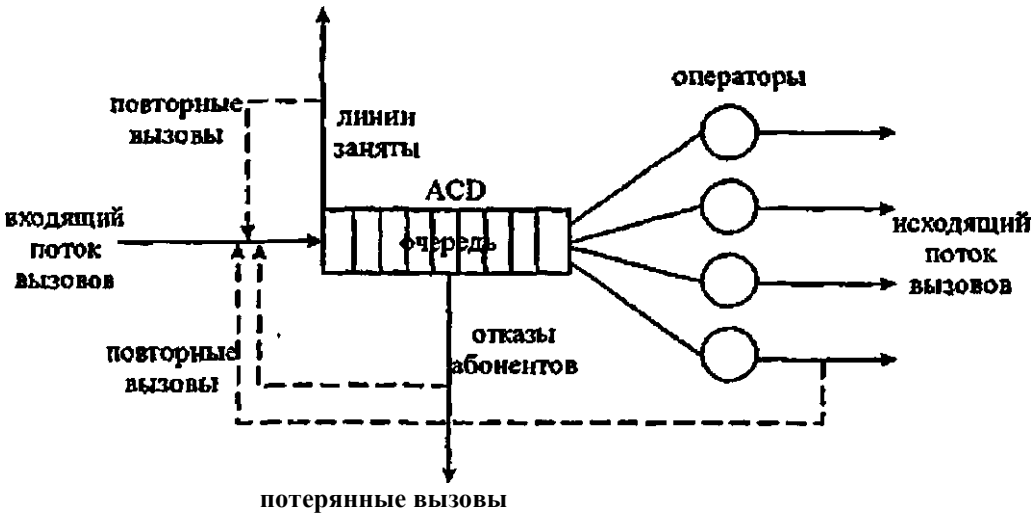


Рисунок 1.2 - Модель центра обслуживания вызовов в виде СМО

Пользователь набирает один из номеров, закрепленных за ЦОВ. Если все входящие линии заняты, звонящий получит отказ в обслуживании (блокировка вызова) и произойдет одно из двух действий: он либо совершит повторный вызов либо не позвонит вовсе, что будет считаться потерянным вызовом. Если хотя бы одна линия свободна, он подключается к ЦОВ и, в частном случае, слышит ответ электронного цифрового автоинформатора (ГУИ). В процессе интерактивного «разговора» с автоинформатором пользователь может получить исчерпывающую информацию и отключиться от ЦОВ.

Однако, зачастую, для получения необходимой информации или услуг, требуется соединение с оператором. В этом случае, в современных центрах обслуживания вызовов, вызов передается на автоматический распределитель вызовов (АСБ), который обладает возможностями маршрутизации звонков на основе множества критериев. Если подходящий оператор не занят и свободен для обслуживания, то данный вызов незамедлительно маршрутизируется на него. Иначе АСБ задерживает вызов до освобождения, требуемого оператора.

В процессе ожидания в очереди может проигрываться музыка, коммерческая или другого вида информация. Пользователь может решить, насколько необходима для него услуга, чтобы ожидать её в очереди. Если она не так важна, он может просто отключиться от центра обслуживания вызовов и попробовать перезвонить ещё раз, или прекратить свои попытки - этот вариант обозначен на рисунке как «отказы пользователей». Все остальные пользователи в итоге получают ответ от оператора.

Наиболее простым способом моделирования такого Call-центра является применение модели СМО типа M/M/m с m рабочими местами операторов и неограниченным числом мест для ожидания. Несмотря на то, что подобная модель не принимает в расчет возможность потери вызовов из-за занятости линий, «нетерпеливости» пользователя, возможность

многоэтапного обслуживания и т.п., она является приемлемым средством оценки характеристик множества простых центров обслуживания вызовов. Для исследования характеристик центров обслуживания вызовов обычно выбираются интервалы времени, на протяжении которых интенсивность поступления вызовов меняется не значительно. Экспериментально доказано, что распределение интервалов между вызовами и времен обслуживания для Са11-центров ТфОП соответствует показательному.

Рассмотрим один из многочисленных примеров применения такой модели СМО для исследования характеристик Са11-центров. Так рассматриваемся модель M/M/m со следующими характеристиками.
Интенсивность поступления вызовов: $\lambda n = \lambda\ ,\ n = 0, 1, 2, \ldots;$

Интенсивность обслуживания: $\mu_n \begin{cases} n\mu, 0 \le n \le m \\ m\mu, n \ge m \end{cases}$

Рассматриваемая система хорошо изучена, для неё известны следующие результаты[11]. Если принять за λ интенсивность поступления вызовов на КЦ, а за μ=1/b среднюю интенсивность обслуживания, то при $A = {}^{\lambda}/_{\mu} = \lambda \cdot b$

$$\rho = {}^{\lambda}/_{\boldsymbol{m}\mu} = {}^{\boldsymbol{A}}/_{\boldsymbol{m}}, \qquad (1.1)$$

где A - поступающая нагрузка,
ρ - коэффициент использования системы.

Для системы M/M/m известно выражение P{W >*0*}=C(т,A)=E2,т(A) которое также называется С-формулой Эрланга

$$E_{2,m}(A) = 1 - \frac{\sum_{k=0}^{m-1} {}^{A^k}/_{k!}}{\left[\sum_{k=0}^{m-1} {}^{A^k}/_{k!} + \left(\frac{A^m}{m!}\right) \cdot \left(\frac{1}{1 - {}^{A}/_{m}}\right)\right]}, \qquad (1.2)$$

где A < m

Среднее время ожидания обслуживания в такой системе вычисляется как

$$W = E_{2,m}(A) \cdot \left(\frac{1}{m}\right) \cdot \left(\frac{1}{\mu}\right) \cdot \left(\frac{1}{1-\rho}\right), \qquad (1.3)$$

На рисунке 1.3 представлена полученная опытным путем диаграмма зависимости между ρ - коэффициентом использования системы и временем ожидания обслуживания вызова простым Call-центре.

Экспериментальные данные близки к получаемым из приведенных выше соотношений. Однако, при всем удобстве и простоте такого подхода, он не учитывает важных особенностей функционирования IP контакт-центрах.

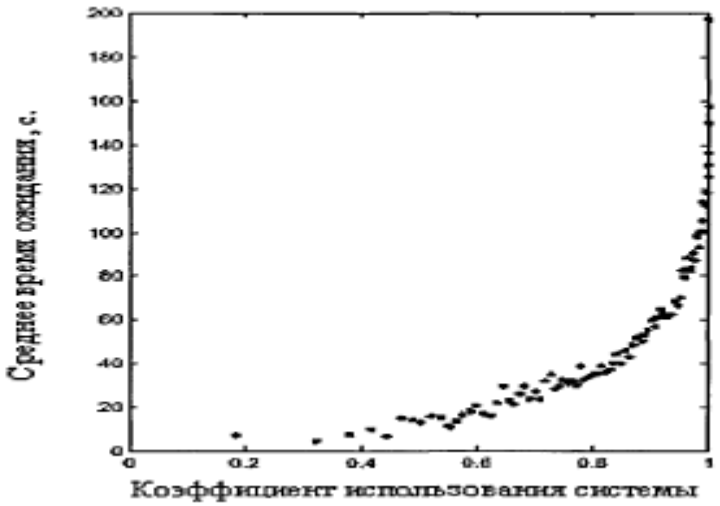


Рисунок 1.3 - Диаграмма зависимости между коэффициентом использования системы и временем ожидания обслуживания вызова Call-центре

В ряде случаев, при исследовании центров обслуживания вызовов, можно столкнуться с произвольным временем обслуживания заявок. В таком случае может быть применена модель СМО вида M/G/n,для неё известно следующее аппроксимационное выражение

$$W \approx W_{M/M//n} \cdot \left(\frac{1+C_v^2}{2}\right), \quad (1.4)$$

где W - среднее время ожидания СМО типа M/G/n, $W_{M/M//n}$ среднее время ожидания СМО типа M/M/n;

$C_v = {\sigma(b)}/{b}$ коэффициент вариации,

$\sigma^2(b) -$ - дисперсия , S - среднее время обслуживания.

В случае большой нагрузки на систему $\left(C(m, A \approx 1)\right)$это выражение принимает вид

$$W = \left(\frac{1}{m}\right) \cdot S \cdot \left(\frac{1}{1-\rho}\right) \cdot \left(\frac{1+C_v^2}{2}\right), \quad (1.5)$$

Данное выражение может применяться для оценки искомых значений, когда получение более точных результатов аналитически затруднено.

Учесть возможность блокировки вызова по причине отсутствия свободных линий позволяет применение модели СМО вида M/M/ m с

отказами, для которой известна В-формула Эрланга, описывающая долю времени, когда все обслуживающие приборы системы заняты. Вероятность занятости всех обслуживающих приборов для такой системы

$$P_N = \frac{\frac{\left(\frac{\lambda}{\mu}\right)^N}{N!}}{\sum_{i=0}^{n} \frac{\left(\frac{\lambda}{\mu}\right)^J}{J!}}, \qquad (1.6)$$

где λ - интенсивность поступления вызовов,

μ - интенсивность обслуживания,

N -число обслуживающих приборов (для данной системы число операторов или входящих линий).

Таким образом, можно узнать основную характеристику системы, прямо влияющую на качество предоставления информационной услуги.

Близкими к оборудованию реальных Call-центров являются модели СМО с ограниченным буферным пространством. Рассмотрим, одну из таких моделей - СМО M/M/m/K.

Отметим, что модель M/M/m/K близка по своим свойствам к рассмотренной выше M/M/m, за исключением ограниченного числа мест для ожидания, при переполнении которого поступающие заявки начинают теряться. Предполагается, что $K \geq m$, т.к. в противном случае некоторые обслуживающие приборы никогда бы не занимались, и система функционировала бы как M/M/m с отказами. Для описываемой системы интенсивность поступления заявок

$$\lambda_n = \lambda,\ n=0,1,\ldots\ldots,k-1$$

Интенсивность обслуживания

$$\mu_n = \begin{cases} n\mu,\ n = 1,2,\ldots m-1 \\ n\mu,\ n = m, m+1, \ldots, k \end{cases}$$

Соотношение, определяющее вероятность заданного числа заявок в системе

$$P_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} \cdot p_0, n = 1,2,\ldots m-1 \\ \frac{\lambda^n}{m!m^{n-m}.\mu^n} \cdot p_0\ , n = m, m+1, \ldots, k \end{cases}, \qquad (1.7)$$

Определяя $\rho = {}^{\lambda}/_{m\mu} =$, получаем

$$\text{P n} = \begin{cases} \frac{(mp)^n}{n!} \cdot p_0, n = 1,2, \dots m-1 \\ \frac{p^n m^n}{m!} \cdot p_0 \, , n = m, m+1, \dots, k \end{cases}$$

Исспользуя известное равенство

$\sum_{n=0}^{k} P_n = 1$, можно найти P_0

Среднее число вызовов в очереди и среднее число вызовов в системе определяется следующими выражениями

$$\text{Mn}_\text{q} = \sum_{\text{n=m+1}}^{\text{K}} (\text{n} - \text{m}) \cdot \text{p}_\text{n}$$

$$\text{Mn} = \sum_{n=1}^{K} n \cdot p_n$$

Известно, что все вызовы, поступающие на систему, когда она находится в состоянии п = К, теряются. Действительная (эффективная) интенсивность поступления заявок в систему вычисляется как

$$\lambda' = \sum_{n=0}^{K-1} \lambda \cdot p_n = \lambda \cdot \sum_{n=0}^{k-1} p_n = \lambda \cdot (1 - p_k), \qquad (1.8)$$

где $p_к$ - вероятность нахождения системы в состоянии К.

Разность λ-λ' = λ- $p_к$ определяет интенсивность потерянных вызовов. В данной модели заявки не могут быть потеряны после поступления в очередь. Воспользуемся формулой Литтла для определения среднего времени ожидания обслуживания

$$\text{W} = \frac{\text{Mn}_\text{q}}{\lambda'} = \frac{\text{Mn}_\text{q}}{\lambda \cdot (1 - p_k)} \, . \qquad (1.9)$$

Для модели M/M/m с неограниченной очередью загрузка системы определяется по формуле

$$\rho = \lambda / m \cdot \mu \, , \qquad (1.10)$$

В случае ограниченного размера очереди она будет равна

$$\text{U} = \lambda' / m \cdot \mu = \rho \cdot (p_k) \qquad (1.11)$$

Однако ни модель СМО M/M/m , ни M/M/m/К не в состоянии учесть возможность ухода вызова из очереди, например, когда пользователь кладет

трубку, не дождавшись обслуживания. Этот недостаток позволяет устранить применение моделей вида M/M/m+M и M/M/m/K+M, где учитывается "терпеливость" пользователя (например, подчиняется экспоненциальному распределению). СМО, M/M/m+M обозначается как "Erlang A"(от англ. - abandonment).

Перечисленные способы исследования характеристик операторских центров, не смотря на свою простоту, хорошо подходят для приблизительной оценки ресурсов необходимых для выполнения задач по обслуживанию вызовов, и нашли свое применение в индустрии call-центров. Тем не менее, они не всегда достаточны для изучения поведения таких сложных систем, как мультисервисные контакт-центры. Математические модели современных центров обслуживания вызовов.

1.3 Математические модели современных центров обслуживания вызовов

В течение всего дня количество вызовов, поступающих на ЦОВ, меняется. Для обслуживания различных вызовов в минимально возможное время требуется определенное количество операторов. К счастью, некоторые типы вызовов не требуют быстрого ответа и могут быть обслужены через некоторое время. Например, телефонные звонки имеют самый высокий приоритет и должны быть обслужены в течение нескольких секунд или минут, ответы на электронную почту или факс могут быть отложены на несколько часов или даже дней. Таким образом, различным типам вызовов можно присвоить различные приоритеты в обслуживании, соответствующие их чувствительности к времени ожидания. Чем меньше должна быть задержка, тем больше приоритет вызова при обслуживании. С учетом различных типов вызовов можно значительно повысить эффективность функционирования ЦОВ путем распределения нагрузки в течение всего дня - в периоды низкой загруженности операторы ЦОВ могут обслуживать низкоприоритетный трафик.

В зависимости от особенностей многофункционального центра обслуживания вызовов для предсказания его поведения могут использоваться различные способы. Так для моделирования простейшего случая обработки вызовов поступающих от ТфОП и сети IP-телефонии можно предложить систему вида M1+M2 / M / c / K1+K2 с входящими пуассоновскими потоками различной интенсивности, одинаковым распределением времени обслуживания и различным числом мест в системе для вызовов от абонентов ТфОП и сети IP- телефонии (подробнее см. [2]).

Определив вероятности нахождения в такой системе Pij заявок обоих типов 0=1,2), зная максимальное число различных заявок в системе Ki и приняв предположение о значительном превосходстве параметра *K2* над *K1*, несложно получить вероятность блокировки для вызовов первого типа (из

ТфОП), а также среднее число вызовов обоих типов в системе и средне время пребывания в системе

$$\text{РВ} = \sum \text{P}_{\text{k}_{\text{i,j}}} . \quad (1.12)$$

Большинство крупных центров обслуживания вызовов, которые и нуждаются в удобных средствах прогнозирования сильнее остальных, работают с несколькими службами и группами операторов. Рассмотрение подобных систем затрудняется тем, что они требуют отдельных решений для многих частных случаев. Если взять частный случай, когда на операторов ЦОВ одновременно поступают 2 пуассоновских потока с различными интенсивностями, причем операторы делятся на 3 группы, первая обслуживает только вызовы первого потока, вторая - обоих, а третья - вызовы только второго потока, то потребуется рассматривать 4 потока: $\lambda_{11}, \lambda_{12}, \lambda_{22}, \lambda_{23}$где i — номер потока, а j - номер группы операторов. Предполагая, что процессы с интенсивностями λ_{ij} -пуассоновские, что может не соответствовать действительности, необходимо будет использовать СМО M/G /m, это позволит учесть различия в интенсивностях обслуживания разных потоков вызовов.

Подводя итоги, 1.2 и 1.3 параграфы можно сказать, что рассмотренные модели СМО, нашли широкое применение для исследования характеристик контакт центров. Однако, результаты, полученные в [11, 12, 13] не позволяют решить задачи диссертационной работы, так как не учитывают особенностей, рассматриваемых в параграфы 1.4 и 1.5. Это является следствием появления новых видов обработки нагрузкой, таких как обработки запросов, в свободной форме, и рассматривающими модели могут быть недостаточно подходящими.

Специфики современные контакт-центров, как объекта исследования излагается в следующем разделе.

1.4 Определение off-line контакт-центра

Типичный алгоритм работы контакт-центра предполагает непосредственное обслуживание запроса при поступлении голосового вызова, когда пользователь информационных услуг задает вопросы оператору центра и получает ответы. В этом случае на скорость работы оператора и типы информационных запросов пользователя, накладывается целый ряд ограничений. По мере развития поисковых систем общего пользования, доступных в сети интернет, растет и уровень требований, которые пользователи предъявляют к информационным услугам контакт-центров.

На практике, современный контакт-центр может обслуживать запросы на самые разнообразные услуги - однотипные справочные запросы, запросы к экстренным службам, запросы на обработку исходящих телемаркетинговых вызовов и прочее. Вместе с тем, в последнее время появляются услуги

контакт- центров, предусматривающие возможность обработки запросов в свободной форме. К таким запросам относятся любые обращения пользователя в контакт- центр, начиная от вопроса по курсам валют на текущий момент до имени президента какого-либо удаленного и небольшого государства. Отличительная черта предоставления подобной информационной услуги состоит в отложенном обслуживании запроса пользователя, а именно: после нахождения ответа, на запрос оператор контакт-центра связывается с пользователем наиболее удобным для него способом, - посредством голосового вызова, через службу SMS или электронную почту.

1.5 Специфические особенности мультисервисного контакт центра, как объекта исследования

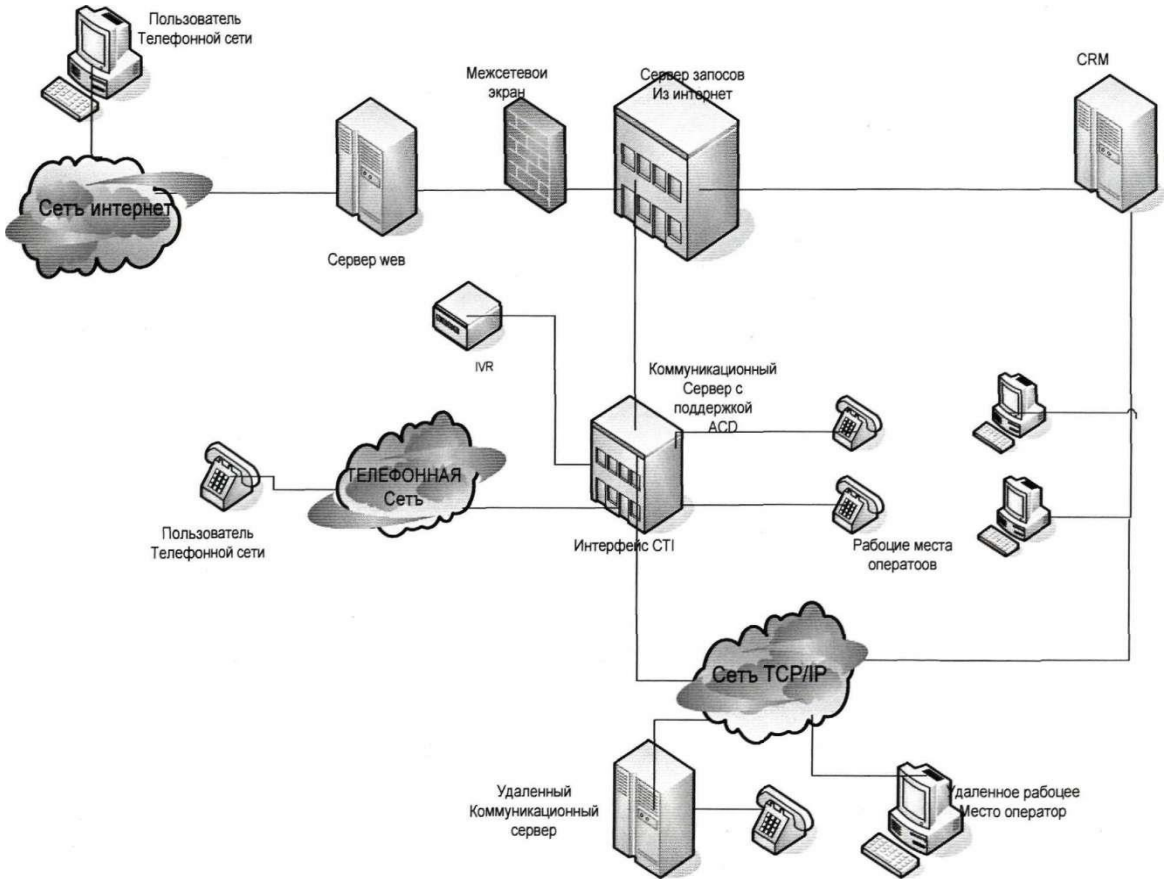


Рисунок 1.4 - Схема построение мультисервисного КЦ

При рассмотрении специфических особенностей контакт-центров, в рамках данной диссертационной работы, основной интерес представляет «среда обитания» объекта исследования. Кратко рассмотрим причину перехода от традиционного для Call-центров окружения к более сложным структурам и выделим отличительные особенности нового вида контакт-центра - МКЦ.

Последние годы происходит интеграция телефонной и компьютерной индустрии, которая привела к появлению так называемых мультисервисных контакт-центров (МКЦ) на базы IP- технологий, обладающих несравненно большим набором услуг и большими возможностями. Контакт-центр нового поколения, структура которого представлена на рисунке1.4, должен обеспечивать прием традиционных телефонных вызовов, поступающих из сетей подвижной связи (СПС), телефонных вызовов, поступающих из сети Интернет с использованием технологий VoIP (skype, yahoo messenger, video), прием заявок по факсу, электронной почте, запросов по технологиям мгновенного обмена сообщениями, прием заявок, допускающих отложенную обработку и запросов из социальные сетей - Вконтакте, Facebook, Twitter. Таким образом, МКЦ обеспечивает обработку и распределение всех видов интернет-запросов, сохраняя возможность распределения телефонных вызовов. Иными словами, обеспечивая совместимость с наиболее распространенными браузерами (MS Internet Explorer, Opera, Firefox). Таким образом, МКЦ обрабатывает множество типов запросов, таких, как: электронная почта; Интернет-чат; обратный вызов (web call back), обращение пользователя через IP- канал связи (web call through VoIP), одновременный просмотр страниц в Интернете оператором и пользователем обработка сообщении из социальные сетей (Вконтакте, Facebook, Twitter.)

Сообщения и запросы в МКЦ обрабатываются в универсальной очереди наряду с телефонными вызовами. В частности Особенность обработки электронной почты связана с использованием интеллектуальных возможностей по анализу и классификации сообщений. Кроме распределения сообщений по операторам МКЦ позволяет реализовать такие функции, как автоматический ответ на сообщения по наиболее подходящим шаблонам, автоматическая отправка подтверждения о получении сообщения.

Компании, внедрившие МКЦ на базе IP-технологий, получают немаловажные преимущества. К таким преимуществам можно отнести:

– независимость размещения - ключевым преимуществом МКЦ на основе IP является его независимое размещение. Так, вне зависимости от физического местоположения оператора, он может выполнять свою обычную работу (принимать и обрабатывать вызовы) по корпоративной сети. Если службы МКЦ расположены в разных местах (разных офисах компании), МКЦ обеспечит интеллектуальную маршрутизацию вызовов независимо от местоположения необходимых ресурсов. С помощью территориально распределенного МКЦ компания имеет возможность полнее задействовать своих сотрудников, находящихся в разных подразделениях, организовать удаленные рабочие места и разрешать работу на дому. Такая гибкость допускает привлечение к работе дополнительного персонала, что позволяет предложить пользователям возможность обращаться в МКЦ в любое время суток. Возможно, наиболее значительная польза от работы сотрудников дома - уменьшение их усталости;

– внедрение и работа объединенной сети - преимущества объединения голоса и данных описывались достаточно много: объединенная сеть позволяет сократить расходы наполовину. Менее известны преимущества МКЦ на основе IP, предлагаемые для работ с заказчиками, хотя только это может стать достаточной причиной для подобного объединения. Поддержка передачи голоса по пакетной сети на базе протокола IP позволяет объединить сети передачи данных и голосовые сети в единую инфраструктуру. Создание и эксплуатация объединенной сети не только дешевле, но и позволяет задать единые правила работы, что гарантирует качественное обслуживание заказчиков. Плюс к этому объединенная сеть, поддерживающая передачу голоса по IP (VoIP), допускает внедрение новых приложений от различных разработчиков и возможность организации новых услуг на основе IP;

– мультимедийные каналы в МКЦ - не менее важной особенностью ЦОВ является поддержка комбинированных каналов связи. При постоянной конкуренции обычной телефонной связи уже недостаточно; для работы с заказчиками требуются и текстовые диалоги-чаты, и электронная почта, и видеосвязь, и возможность совместной работы в Web. Поскольку все перечисленные возможности реализованы на открытых стандартах, они безболезненно интегрируются в открытую архитектуру МКЦ, использующего протокол IP, и ими можно управлять как составной частью единой системы работы с заказчиками. Участники рынка понимают, что для сохранения конкурентоспособности необходимо управлять работой со всеми заказчиками централизованно - с помощью МКЦ - и постепенно переходить к индивидуальной работе с каждым заказчиком;

– быстрое внедрение новых приложений - еще одно важное преимущество МКЦ заключается в поддержке быстрого внедрения новейших приложений, причем внедрения более скорого, чем обычные. Поскольку работа ведется в объединенной IP-сети, приложения не зависят от операционных систем - при том, что их совместимость с другими IP-приложениями гарантируется;

– стратегия перехода от традиционной телефонии к IP-технологиям - любая новая технология, появляющаяся на сложившемся рынке, обязательно должна поддерживать обратную совместимость с более ранними технологиями в своей области. Это особенно верно для МКЦ, в техническую базу которых многие годы вкладывались большие средства. Новые технологические решения должны интегрироваться с имеющимся и обеспечивать плавный и безболезненный переход на новый этап развития. Такая стратегия не только сокращает простои, но и позволяет вести внедрение постепенно, давая персоналу время освоить новые возможности технологии. МКЦ предлагаю ттакую стратегию перехода, при которой IP-технологии некоторое время сосуществуют с традиционной телефонией. Решение МКЦ сводит эти две несопоставимые технологии в объединенную систему, где можно просматривать текущие и хронологические отчеты, а операторов, работающих в сети IP, и операторов, работающих в сети традиционной

телефонии, можно разбить на тематические группы. Независимо от IP- или традиционной связи и те, и другие операторы одинаково принимают поступающие вызовы, а управление и отчетность ведутся одинаковым образом.

Описание модели МКЦ начинается с описания основных сервисов, которые он будет обеспечивать и наиболее популярные услуги являются:

– организация «горячих» линий;

– опросы, анкетирование, социологические и маркетинговые исследования;

– поддержка рекламных кампаний и определение эффективности рекламы;

– приём и поддержка заказов на продукцию и услуги от лица заказчика;

– организация телефонных конференций;

– проведение массовых телефонных оповещений (в том числе оповещение клиентов заказчика об оплате и состоянии счетов, о задолженностях по оплате товаров и услуг);

– создание и ведение баз данных заказчиков;

– продажа товаров и услуг заказчика через глобальную информационную сеть и по каталогам;

– до и после продажная поддержка клиентов заказчика;

– организация телефонных голосований, телефонных лотерей, конкурсов.

Актуальными и определяющими цель исследований данной диссертационной работы является появления новейших услуги контакт-центров, предусматривающие возможность обработки запросов в свободной форме, таких контакт-центров называются off-line контакт-центр.

Главная, в контексте данной диссертационной работы, особенность МКЦ, по сравнению с предшествующими центрами - это способность обслуживать вызовы нескольких классов, поступающих из разных телекоммуникационных сетей:

– запросы речевой связи - из ТфОП;

– запросы речевой связи - из СПС;

– запросы речевой связи - из Интернета, с использованием технологии IP- телефонии - Skype, Yahoo messenger;

– запросы связи по факсу, электронной почте;

– запросы связи в режиме текстового чата - из Интернет, СПС;

– запросы из социальные сетей - Вконтакте, Facebook, Twitter;

– запросы с отложенным обслуживаниям.

Определение ВВХ в основных подсистем такого контакт-центра является весьма актуальной и новой задачей, а к тому же не тривиальной, так как необходимо учитывать различную интенсивность поступления вызовов из ТфОП, интернет, по электронной почте, факсимильные вызовы и т.д. Также, для разных типов запросов, имеют место законы распределения интервалов

поступления и времени обслуживания, крайне отличные от принятых для исследования старых систем, и возможность сложного распределения вызовов разных типов по группам операторов. Для каждого типа вызовов может быть своя максимально допустимая длина очереди, различные приоритеты и допустимое время ожидания обслуживания.

Разнообразие типов обслуживаемых запросов приводит к существенным изменениям функциональной структуры рассматриваемого IP- контакт-центра, по сравнению с системами прошлого поколения.

В случае контакт-центра запросы поступают с разной интенсивностью от источников разного типа, допускают разную длительность ожидания и разную продолжительность и законы обслуживания, т.е. различаются параметрами, которые определяют характеристики входящей нагрузки и на основании которых обычно производится распределение вызовов и организация очередей. Так, например, пороги длительности ожидания для заявок, допускающих отложенное обслуживание, могут измеряться десятками минут, а для традиционных телефонных вызовов - десятками секунд. Фактически, очередь ожидания превращается в буфер, выбор заявок из которого производится не в порядке их поступления, а на основе анализа нескольких параметров, характеризующих эти заявки.

Механизмы обслуживания разных заявок могут быть различными. Их могут обслуживать либо отдельные операторы или группы, либо одни и же операторы. Таким образом, можно выделить две основных, в контексте данной работы, особенности контакт-центров, по сравнению с предшествующими им системами. Обе они приводят к необходимости создания нового подхода к анализу такого рода систем.

Очевидно, что для исследования контакт-центров, обслуживающих не только голосовые вызовы, удобным является использование механизма приоритетов. Так, система, обрабатывающая вызовы из ТфОП, сети IP-телефонии, а также электронную почту и текстовые запросы Instant Messaging, могла бы моделироваться СМО со смешанной дисциплиной приоритетов, абсолютной или относительной (зависящей от времени ожидания заявки). Вторая особенность изучаемых систем состоит в том, что рассматриваемый контакт-центр позволит обслуживать как прямые (on-line), так и отложенные (off-line) запросы, причем со всех видов обращений включая телефонный вызов. Такие контакт-центры (КЦ) можно назвать, on/off-line КЦ, и это есть объект исследования диссертации. Как показывают недавние экспериментальные исследования [16, 17], здесь практически не применимы старые методики расчета телекоммуникационных систем с входящими пуассоновскими потоками и показательным распределением времени обслуживания, и необходимо рассмотрение процессов с самоподобными свойствами. На рисунке 1.5 представлена иллюстрация разных типов запросов, поступающих в МКЦ.

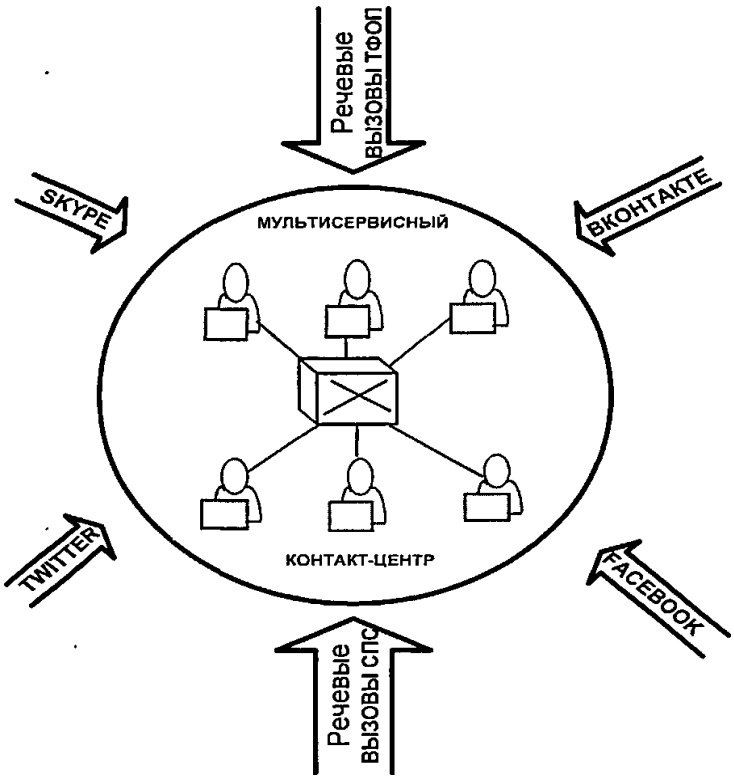


Рисунок 1.5 - Иллюстрация разных типов запросов поступающих в МКЦ

2 Приоритетные модели обслуживания заявок в мультисервисном контакт-центре

2.1 Функциональная модель разноприоритетного трафика операторской подсистемы МКЦ

Продолжим начатое в главе 1 рассмотрение современных МКЦ, построенных на базе IP-технологий. Формализируя качественное описание, предложенное в параграфе 1.5, введем в рассмотрение общую модель операторской подсистемы (рисунок 2.1). Операторская подсистема МКЦ, являющаяся одним из основных элементов обслуживания запросов пользователей на информационные услуги, которые требуют участия оператора. Данная подсистема обрабатывает речевые и факсимильные запросы, приходящие из ТфОП, СПС, сетей IP-телефонии, а так же текстовые запросы пользователей электронной почты, СПС, систем интерактивного обмена текстовыми сообщениями и также пользователей социальных сетей. Выделим компоненты, входящие в её состав:

– множество *накопителей заявок* (очередей), необходимых для сглаживания всплесков нагрузки, поступающей от разных сетей;

– *обслуживающие приборы* (терминалы операторов-агентов), принимающие запросы из накопителей заявок в соответствии с ё приоритетной дисциплиной обслуживания и в порядке освобождения. Исходя из общего описания контакт-центра, приведенного в главе 1, запросы поступают на операторскую систему от множества источников по соответствующим им инфокоммуникационным сетям, таким, как ТфОП, СПС, сети IP-телефонии, Интернет.

В отличие от центров обслуживания вызовов ТфОП, операторская подсистема IP-контакт-центр должна взаимодействовать со многими разнородными источниками запросов на предоставление информационных услуг, вследствие чего возникает сложная организация накопителя заявок и многофункциональные обслуживающие приборы.

Интенсивность поступления заявок и не терять их при умеренных всплесках нагрузки. В зависимости от применяемой дисциплины обслуживания и числа различных классов заявок, организация накопителей заявок контакт центров может различаться от простой очереди до динамических систем с относительными приоритетами. В этом случае аналитическое рассмотрение операторской подсистемы МКЦ усложняется, однако, как будет показано позже, остается возможным и может иметь практическое применение.

В отличие от систем предыдущих поколений, операторская подсистема контакт-центра имеет более сложными, многофункциональными терминалами рабочих мест операторов. Это позволяет обходиться при моделировании однотипными обслуживающими приборами.

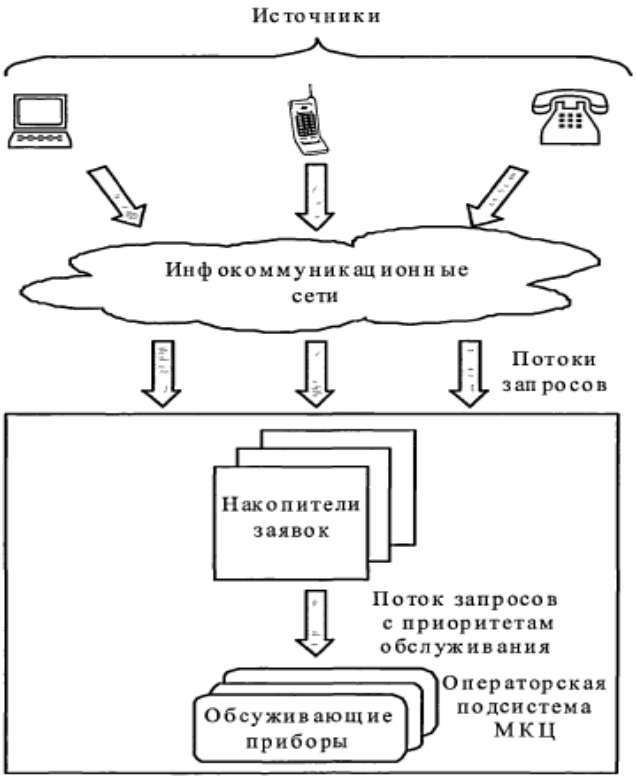


Рисунок 2.1 - Функциональная модель операторской подсистемы МКЦ

Как уже отмечалось, назначение накопителя заявок операторской подсистемы МКЦ - компенсация резких всплесков поступающей нагрузки. Работа долговременными изменениями нагрузки сопровождается адаптацией под эти изменения набора обслуживающих приборов, например, путем их добавления или отключения. Кроме достижения требуемых временных рамок при предоставлении информационных услуг, такой подход может решить задачу минимизации затрат на обеспечение функционировании подсистемы.

Если решение первой проблемы при предоставлении услуг (кратковременные всплески) заключается в применении адекватных средств проектирования операторской подсистемы, то вторая проблема требует предоставления управляющему персоналу контакт-центра соответствующего по прогнозирования поведения ВВХ подсистемы в зависимости от поступающей нагрузки. Средства проектирования и прогнозирования в обоих случаях будут иметь в своей основе одну и ту же модель операторской подсистемы контакт- центра.

Так как существование всплесков и плавных изменений нагрузки, очевидно, то с учетом широкого диапазона характеристик источников заявок необходимо ввести такое управление процессом удержания заявок в накопителях операторской подсистемы, чтобы отрицательные последствия этих событий были минимальными.

Известно, что характеристики трафика, поступающего от источников заявок, могут меняться во времени. В зависимости от типов

инфокоммуникационных сетей, и вида предоставляемых контакт-центром услуг возникают периодические возрастания и спады интенсивностей поступающих запросов. Это могут быть как кратко и долговременные всплески интенсивности поступления, которые являются следствием случайного её характера. При этом, в случае МКЦ, имеет место большое число разных типов поступающих запросов на информационные услуги, причем, кроме чисто технических отличий в средствах их транспорта, наблюдаются существенные отличия их ВВХ [17, 19, 20]. Данная особенность отсутствовала в эпоху старых ЦОВ и традиционной телефонной нагрузки, но исследование ВВХ контакт- центров требует, прежде всего, учета новых особенностей поступающего трафика.

2.2 Анализ трафика, поступающего в МКЦ, и процессов его обслуживания

Определим, какими интерфейсами обладает МКЦ. Далее, отталкиваясь от технических особенностей и экспериментальных данных, найдем аналитическое описание ВВХ потоков трафика каждого интерфейса.

В рассматриваемой в диссертационной работе системе предоставления информационных услуг раздельного внимания требуют операторская подсистема МКЦ. Как уже отмечалось, в общем случае она осуществляет обслуживание запросов на информационные услуги приходящие:

– в речевом виде от ТфОП, СПС, сетей IP-телефонии;

– в текстовом виде от систем интерактивного (диалогового) обмена сообщениями сети Интернет, СПС и социальны сетей;

– в текстовом виде от систем обмена сообщениями позволяющих отложенную обработку ТфОП, СПС, сетей 1Р-телефонии и Интернет (факсимильные сообщения и электронная почта);

– в речевом и текстовом виде с отложенной обработкой.

Следовательно, интерфейсы операторской подсистемы МКЦ могут быть разделены на следующие группы: речевой, текстовый диалоговый и речевой - текстовый с отложенной обработкой. Потоки запросов на информационные услуги, приходящие на каждую из групп интерфейсов, могут отличаться в своих ВВХ.

Выделим основные в контексте данной работы особенности процессов поступления и обработки потоков запросов по указанным группам интерфейсов. Для речевой группы интерфейсов особенности ВВХ поступающих потоков известны ещё из базовых работ по теории телетрафика, рассматривавших процессы, происходящие в ТфОП. Для центров обслуживания вызовов ТфОП эти данные приводятся в работах [7, 8, 9, 10].

Установлено, что потоки речевых вызовов, приходящие на операторскую подсистему от большого числа источников, имеют показательное распределение интервалов времени между поступающими запросами и аналогичное распределение времени обслуживания. Данный факт

имеет экспериментальные подтверждения во множестве классических работ по обслуживанию телефонной нагрузки.

Стоит отметить, что в некоторых работах, например [11], авторы пытаются исследовать речевой трафик телефонных сетей, применяя теорию самоподобных процессов, без чего, как показывают современные исследования, сложно обойтись при анализе процессов, происходящих на всех уровнях сети Интернет. Однако верность применения при анализе телефонной речевой или речевой нагрузки показательных законов распределения многократно доказана.

Кроме речевых интерфейсов, операторская подсистема содержит также два типа текстовых. Технической реализацией текстового диалогового интерфейса и текстового интерфейса с отложенной обработкой, например, могут быть системы интерактивного обмена текстовыми сообщениями Web chat и электронная почта интернет.

Согласно экспериментальным исследованиям, приведенным в [21], закон (таблица 2.1), по которому происходит поступление запросов установления сессий обмена информацией на прикладном уровне сети Интернет, соответствует показательному распределению интервалов времени между запросами. То же указано в [22] и [23].

Зависимости для процессов обслуживания запросов, поступающих через текстовые интерфейсы операторской подсистемы контакт-центра, могут заметно отличаться от привычных для расчетов телефонной нагрузки. Они могут являться медленно-затухающими распределениями [20, 22, 23, 24, 25], иначе называемыми распределениями с «тяжелым хвостом» (heavy-tailed). В первую очередь, это проявляется в наличии у рассматриваемых процессов больших значений дисперсии.

Как уже отмечалось, назначение накопителя заявок операторской подсистемы МКЦ - компенсация резких всплесков поступающей нагрузки. Работа долговременными изменениями нагрузки сопровождается адаптацией под эти изменения набора обслуживающих приборов, например, путем их добавления или отключения. Кроме достижения требуемых временных рамок при предоставлении информационных услуг, такой подход может решить задачу минимизации затрат на обеспечение функционировании подсистемы

В качестве одного из основных положений при исследовании модели контакт-центра используется допущение, что все поступающие заявки имеют длительности, распределенные по показательному закону, в том числе и длительности текстовых запросов (e-mail, SMS, MMS, chat...). Указанные административные ограничения являются принятой практикой и распространяются на все типы запросов: речевые, текстовые, мультимедийные.

Указанное свойство серьезно влияет на производительность оборудования, что заставляет учитывать его при проведении аналитического и имитационного моделирования. Но в реальных контакт-центр закон

обслуживания заявок может быть лишен heavy-tailed распределений благодаря административным ограничениям.

Т а б л и ц а 2.1- Законы распределения случайных величин Web-служб

Величина	Тип распределения
Обьем информации, передаваемый в течение сессий	Логнормальное, Парето
Продолжительность сессий	Логнормальное, Парето
Размеры запрашиваемых с серверов файлов	Логнормальное, Парето или смешанное
Периоды поступления запросов на организацию сессии	Показательное

В качестве обоснования принятого допущения можно принять ориентацию оборудования и организации МКЦ на обслуживания массовых запросов имеющих однотипный характер.

Таким образом, поступающие на подсистему потоки могут вполне достоверно моделироваться показательным законом распределения времени между поступающими запросами. Этот же закон распределения хорошо моделирует процессы обслуживания запросов, приходящих через группу речевых и текстовые интерфейсов МКЦ. Более подробно ВВХ операторской подсистемы рассматриваются в следующем параграфе.

2.3 Подходы к исследованию ВВХ операторской подсистемы мультисервисных контакт-центров

В разделе 1.2 рассматривались подходы к моделированию традиционные ЦОВ. Перед исследованием ВВХ операторской подсистемы контакт-центров отметим уже предлагавшиеся подходы. Так, в [14], рассматриваются методы расчетов производительности ЦОВ, использующих IP-технологии. Исследуется система, в которую поступают 2 типа трафика: из ТфОП и сети 1Р-телефонии. Поступающие потоки имеют показательные распределения интервалов времени между вызовами, допускается различная средняя интенсивность поступления. Среднее время обслуживания заявок разных типов установлено равным.

В основу модели рассматриваемой системы авторами положена модель СМО M/M/n/K, с числом обслуживающих приборов (операторов) *n* и количеством мест для ожидания *K*, рассмотренная в первой главе диссертации. При проведении математического моделирования вводится предположение о значительном потенциальном превосходстве числа мест для

ожидания, доступным вызовам, поступающим из сети IP-телефонии перед числом мест для ожидания, доступных для вызовов из ТфОП.

Результаты из [14] позволяют предсказать поведение операторской подсистемы контакт-центра, обслуживающей два потока вызовов, имеющих разные средние интенсивности поступления и одинаковое время обслуживания запросов. Данная модель может быть использована также при исследовании контакт-центра, обслуживающего речевые вызовы и запросы электронной почты. Кроме того, авторы рассматривают вопрос планирования канальной емкости для обслуживания заявок из IP-сети.

В [23] рассматривается системы положена модель СМО M/G/n *c* неограниченным числом мест для ожидания, обеспечивающая возможность исследования систем при общем законе распределения времени обслуживания запросов.

Более сложная система рассматривается в [9] и [10]. Исследуется операторская подсистема, на которую с различными интенсивностями поступают два пуассоновских потока вызовов. Операторы делятся на три группы, первая обслуживает только вызовы первого потока, вторая - обоих, третья - вызовы только второго потока. Допускаются различные интенсивности обслуживания вызовов двух потоков.

Авторами разработана модель для определения времени ожидания заявки в очереди при обслуживании вызовов несколькими группами операторов. Предложенный механизм может быть использован при изучении характеристик простейшего контакт-центра, однако не является универсальными и, в зависимости от ситуации, может потребовать кардинального пересмотра.

Меньшая часть работ по исследованию характеристик операторских подсистем посвящена рассмотрению процессов обслуживания вызовов, поступающих в СМО с использованием дисциплин обслуживания с приоритетами. Эти вопросы рассматриваются, например, в [13] и [15]. Так, используя развитый аппарат работы [13], можно осуществить расчет ВВХ для случая, когда на систему поступают голосовые вызовы, требующие скорейшего обслуживания и заявки по электронной почте, позволяющие отложенное обслуживание. Однако эти результаты могут использоваться лишь при исследовании простейших вариантов операторских подсистем контакт-центров. Не смотря на эти технические ограничения, в последней время предлагается модели для определения характеристик более сложных контакт-центрые таких как ЦОВ с аутсорсингом, off-line ЦОВ и т.д.

2.4 Моделирование операторской подсистемы МКЦ

Введем следующие допущения. Заявки на предоставление информационных услуг приходят от источников через случайные интервалы времени. Запросы распределяются по операторам равномерно.

Введение новых услуг способствует более эффективной работе операторов, что увеличивает общую производительность и добавляет гибкость в управлении МКЦ. Происходит это за счёт равномерного распределения нагрузки в течение всего рабочего времени. В периоды низкой занятости основной работой - обслуживанием телефонных вызовов, операторы МКЦ могут обрабатывать трафик пакетных сетей. Однако простая реализация такого подход вносит свои затруднения - ведь телефонные вызовы имеют самый высокий приоритет и должны быть обслужены в течение нескольких секунд или минут. В противном случае будет расти количество отказов от обслуживания, что может привести к потере клиентов. Между тем, ответы на электронную почту, факс или текстовый чат на некоторое время могут быть отложены. Данная проблема известна в западной литературе как интеграция вызовов (call blending). Избежать этой проблема можно путем применения соответствующей дисциплиной обслуживания.

Таким образом, основная цель функционирования группы операторов подсистемы МКЦ состоит в обслуживании речевых вызовов, приходящих от традиционных телефонных сетей, сетей IP-телефонии (Skype, yahoo messenger,video) и подвижной связи; Прием заявок, допускающих отложенную обработку, запросов из социальных сетей - Вконтакте, Facebook, Twitter, а также обработка запросов поступающих в текстовом виде, от пользователей ТфОП, сети Интернет и сетей подвижной связи. К последним относятся запросы средствами факсимиле, электронной почты, средствами систем интерактивного обмена текстовыми сообщениями (IM) и служб коротких сообщений (SMS,MMS).

Наличие такого набора задач вызывает необходимость их иерархического упорядочивания по степени важности и срочности для эффективной работы IP- контакт-центра в целом. В работах по теории телетрафика показано [18], что, если в систему поступают потоки неоднородных запросов, различающихся по относительной важности и длительности обслуживания, то функционирование системы в целом может быть улучшено за счет введения приоритетных дисциплин обслуживания, определяющих в какой последовательности, когда и какой запрос поступает на обслуживание.

Несмотря на больший спектр, предлагаемый информационных услуг, на сегодняшний день, наиболее популярные которые применяют в МКЦ, являются телефония, электронная почта, web-запросы и текстовый чат.

С позиций теории телетрафика исследуемую систему можно рассматривать как многоканальную систему массового обслуживания (СМО) с несколькими классами вызовов и разными приоритетами обслуживания заявок.

Наиболее распространенная мера качества обслуживания пользователей доля обслуженных вызовов, которые ждали в очереди меньше определенного количества времени, $P\{W < T, Sr\}$, где W - время ожидания в стационарном режиме, {Sr} - событие обслуживания пользователя оператором, T- заданное

время, определенное менеджерами ЦОВ. Наиболее распространенным является правило 80/20 - по меньшей мере 80% пользователей должны ожидать в очереди не более 20 секунд, т.е. $P\{W < 20, Sr > \}0.8$.

Другая характеристика качества обслуживания, которая реже используется на практике, P{W>0} - доля вызовов, попавших в очередь. Во многих работах данная характеристика используется для определения режима, в котором функционирует ЦОВ. Обычно, делают аппроксимацию для больших ЦОВ с высокой интенсивностью поступающие нагрузки. Рассмотрим для этого три режима функционирования ЦОВ:

а) Режим качества (quality driven,QD)

$$P\{w>0\} \to 1$$

Данный режим должен использоваться в ЦОВ, где качество обслуживания встает на первый план и превалирует над эффективностью функционирования центра в целом. Пример таких ЦОВ - экстренные службы, ЦОВ с высокодоходными (VIP) клиентами и др.

б) Режим эффективности (efficiency driven,ED)

$$P\{W>0\}$$

Основная ставка делается на эффективность ЦОВ, т.е. на высокую загруженность операторов.

В этом случае фактически все пользователи попадают в очередь и им приходится ждать перед тем, как освободится один из оператора

Данный режим должен использоваться только когда эффективность функционирования ЦОВ встает на первый план. Несколько работ [28,29] действительно показывают, что данный режим, может быть, приемлем для многих ЦОВ, особенно для тех, которые работают не ради получения прибыли. Предельный анализ данного режима доступен в [28,30].

в) Режим баланса между качеством и эффективностью (quality and efficiency driven, QED)

$$P\{W>0\} \to a,\ 0 < a < 1.$$

Режим QED было уделено много внимания в последние несколько лет, особенно I- модель, соответствует нескольким независимым очередей, каждая со своей собственной группой операторов (без дублирования в квалификации). Из-за своих особенностей режима QED, обладало недавно значительным внимание в литература. Все же режим был явно признан уже в 1923 Эрланга (это появилось в [24], который обращается и к Erlang-B (M/M/n/n) и к Erlang-C (M/M/n) модели. Позднее, обширная работа проводилась в различных телекоммуникационных компаний в [12].

Приводится исследование функционирования ЦОВ в данном режиме с экономической точки зрения. Здесь рассматриваются эксплутационные затраты на работу операторов и затраты из-за плохого обслуживания (недополученный доход). Наиболее подходящий расчет штата ЦОВ подчиняется следующему правилу: если *A* обозначает поступающую нагрузку, тогда

$$N \sim A + \beta \cdot \sqrt{A}, \qquad (2.1)$$

где β- параметр качества обслуживания, чем он больше, тем выше уровень обслуживания пользователей.

В режиме QED вероятность ожидания *P{W>0}* является функцией качества обслуживания и отношения ^ (средняя терпеливость, выраженная в единицах среднего времени обслуживания).

На рисунке 2.2 показан зависимость между β и P{W> 0} для изменяющегося соотношения $\frac{\mu}{\theta}$ дополнение, приведен график для модели Эрланг-С, которая рассчитывается только при положительных значениях β. Из рисунка **видно, что для** больших значений соотношения μ/θ (очень терпеливые пользователи), кривые для моделей Эрланг – А и Эрланг –С сходятся ближе.

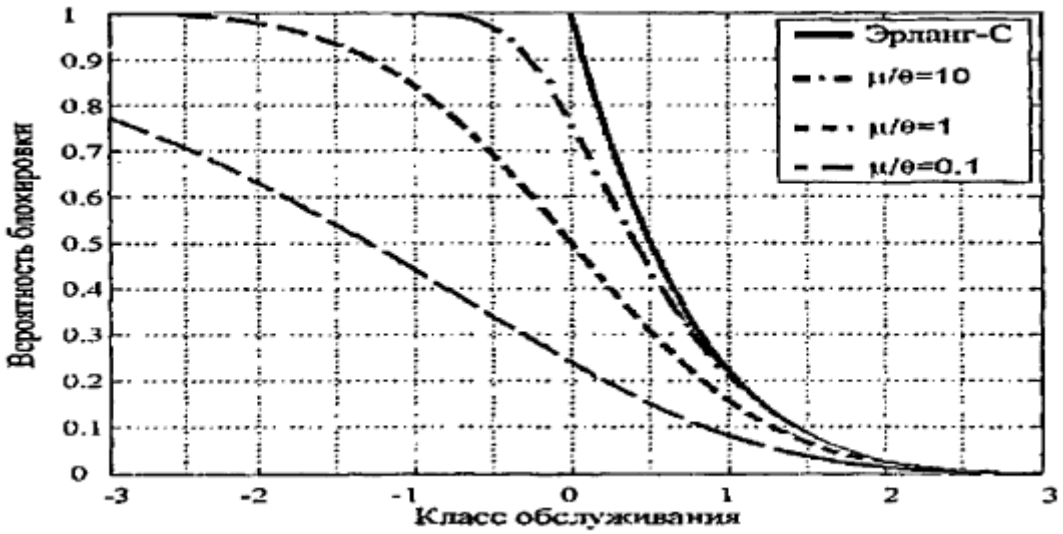


Рисунок 2.2 - Асимптотические соотношения между вероятностью блокировки и качеством обслуживания β

Режим QED позволяет найти золотую середину между загруженностью операторов (близкой к 100%) и качеством предоставляемого сервиса. Параметр качества обслуживания β вычисляется как: $\beta = \frac{N-A}{\sqrt{A}}$

В настоящее время наблюдается бурный рост компаний, предоставляющих современные услуги. МКЦ используются для предоставления услуг по продажам товаров, обслуживания пользователей компании и других специальных транзакций. Как упоминалось, традиционные модели уже не могут отражать все бизнес-процессы, происходящие в современных компаний. Наиболее интересная на сегодняшний день является модель с несколькими классами вызовов и большим количеством операторов, обслуживающими эти вызовы. Эта модель как нельзя лучше отражает картину современных МКЦ. Наиболее перспективными представляются исследования для больших МКЦ с высокой загруженностью операторов. При таких условиях как уже отмечали выше, соблюдается строгий баланс между качеством функционирования МКЦ и загруженностью работающих операторов (режим QED). При рассмотрении такой модели естественно задаться вопросами: Какое количество операторов потребуется для обслуживания поступающих вызовов различных классов и каким образом маршрутизировать вызовы с целью снижения затрат или повышения прибыли с условием сохранения качества обслуживания (QoS) В [18] найдены некоторые важные свойства такого ЦОВ.

Расчет количества операторов мультисервисного ЦОВ ничем не отличается от расчета операторов для ЦОВ с одним классом вызовов при условии одинакового суммарного объема требований. Кроме того, дифференциация уровня обслуживания различных вызовов достигается путем использования простого порогово-приоритетного управления (threshold-priority control).

В режиме QED расчет операторов по правилу квадратного корня где А - нагрузка на ЦОВ, а /5 - постоянная величина) и порогово-приоритетное управление асимптотически оптимальны с соблюдением различным предположений в модели.

Расчет количества операторов и управление работой ЦОВ происходят в различных временных рамках. Если управление происходит в реальном масштабе времени, то планирование штата ЦОВ происходит обычно за неделю. Это само собой предполагает, что расчет количества операторов происходит на основе прогноза нагрузки на ЦОВ. Поэтому при расчете штата, чтобы избежать пересчета или недочета операторов, стараются полагаться на методы, которые требуют только ограниченной информации о будущей нагрузке. Проблеме расчета операторов посвящено множество работ [25,26,27]. Но на сам дело, для расчета общего штата ЦОВ достаточно информации лишь о суммарном количестве требований, вместо прогноза по каждому классу вызовов.

Динамическое управление ЦОВ основано на приоритетах и пороговых уровнях - вызов с определенным приоритетом может маршрутизироваться на обслуживание только в том случае, если в очереди ожидания нет вызовов с более высоким приоритетом, а количество свободных операторов превышает

пороговый уровень для данного класса. Далее, рассмотрим этот метод управления.

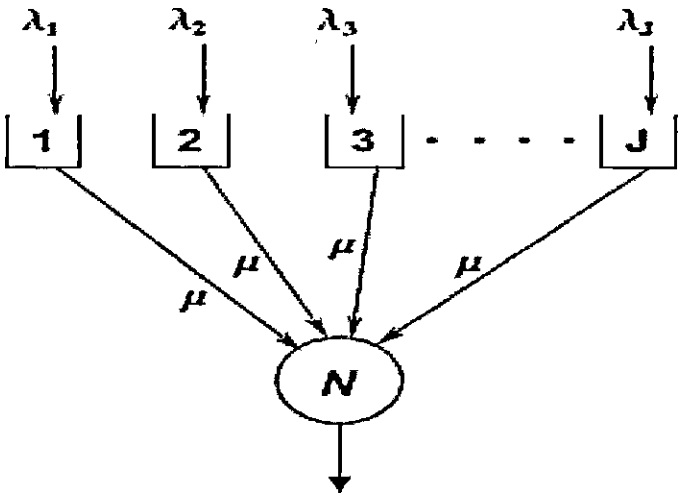


Рисунок 2.3 - Модель МЦОВ с несколькими классами вызовов

Рассмотрим модель, изображенную на рисунке 2.3. В такой модели обслуживаются J классов вызовов. Вызовы поступают согласно

Пуассоновскому закону с интенсивностью λ' для i-го класса. Длительности обслуживания предполагаются распределенными по экспоненциальному закону, а интенсивность обслуживания μ для всех классов вызовов. Дисциплин обслуживания предлагает быть с относительным приоритетом. При поступлении вызова класса i маршрутизировать его к свободному оператору (дисциплина РСРБ) только в том случае, если:

а) очередь j пуста для всех вызовов более высоких классов j (то есть j < i);

б) количество свободных операторов превышает Ki.Здесь 0= K1<K2... < K-пороговые уровни.

Такая модель обычно обозначается как M/M/{Ki }.

Пусть ЦОВ функционирует в режиме QED. Таким образом, количество операторов рассчитывается по формуле $N \sim A + \beta \cdot \sqrt{A}$

Точный анализ модели, в том числе вероятность задержки для каждого приоритетного типа, а также преобразование Лапласа для их ожидания, было проведено в работе [16]. Рекурсивные уравнения, полученные в [16] переводят на весьма сложное выражение даже в случае двух классов вызовов.

Предположим, что оба класса вызовов - низкоприоритетный и высокоприоритетный имеют общее экспоненциально-распределенное время обслуживания, с интенсивностью обслуживания μ. Интенсивность поступления высокоприоритетных вызовов и низкоприоритетных вызовов соответственно λ_1, λ_2 (общая интенсивность λ). Единственное ограничение накладывается на интенсивность λ_2 - она сопоставима с λ, т.е. если λ

стремится к бесконечности, отношение $\frac{\lambda_2}{\lambda}\to\ a_2, a_2 < 2$. Пусть ЦОВ функционирует в режиме QED. Таким образом, количество операторов рассчитывается по формуле

$$N = A + \beta\sqrt{A}, \tag{2.2}$$

где $A = (\lambda_1 + \lambda_2)/\mu$.

Пусть P_n определяет стационарную вероятность того, что занято n операторов (0<n<N). Пусть K - пороговый уровень для низкоприоритетных вызовов (то есть, низкоприоритетные вызовы будут допускаться к обслуживанию, если свободно более чем K операторов). Пусть $M = n - K$. Условие устойчивости выглядит следующим образом

$$\frac{\lambda_1}{N\mu}<1 \ \ и \ \ \lambda_2 h_1(M)<1$$

$$h_1(M) = \frac{1}{M\mu} + \sum_{k=2}^{n-M} \frac{\lambda_1^{k-1}}{\mu^k \prod_{j=0}^{k-1}(M+j)} + \frac{\lambda_1^{n-M}}{(n\mu-\lambda)\mu^{n-M}\prod_{l=1}^{n-M}(n-l)} \tag{2.3}$$

Учитывая эти условия и свойство Пуассоновского потока входящих вызовов PASTA (Poisson arrivals see time averages) , вероятность ожидания для высокоприоритетных вызовов будет определяться выражением

$$P_N = P_0 \cdot \left(\frac{\lambda_1+\lambda_2}{2}\right)^M \cdot \frac{\lambda_1^n}{\mu}\cdot\frac{1}{n!}\cdot\frac{1}{1-\frac{\lambda_1}{n\mu}}\cdot\frac{1}{1-\lambda_2 h_1(M)}, \tag{2.4}$$

где

$$P_0 = \left(\sum_{n=0}^{M-1}\left(\frac{\lambda_1+\lambda_2}{\mu}\right)^n \cdot\frac{1}{n!} + \sum_{n=M}^{N-1}\left(\frac{\lambda_1+\lambda_2}{2}\right)^M \cdot\left(\frac{\lambda_1}{\mu}\right)^n \frac{1}{n!}\cdot\frac{1}{1-\lambda_2 h_1(M)} + \right.$$
$$\left. + \left(\frac{\lambda_1+\lambda_2}{2}\right)^M \left(\frac{\lambda_1}{\mu}\right)^N \left(\frac{1}{N!}\right)\cdot\frac{1}{1-\frac{\lambda_1}{N\mu}}\cdot\frac{1}{1-\lambda_2 h_1(M)}\right)^{-1}$$

Пусть P_{M+} - вероятность ожидания для низкоприоритетные вызовы. Тогда, с условием PASTA , P_{M+}равно вероятность того, что M или более операторы занят

$$P_{M+} = P_0\left(\sum_{n=M}^{N-1}\left(\frac{\lambda_1+\lambda_2}{\mu}\right)^M\left(\frac{\lambda_1}{\mu}\right)^n\frac{1}{n!}\right)\frac{1}{1-\lambda_2 h_1(M)} + \left(\frac{\lambda_1+\lambda_2}{2}\right)^M\left(\frac{\lambda_1}{\mu}\right)^N\cdot\frac{1}{N!}\cdot\frac{1}{1-\frac{\lambda_1}{N\mu}}\cdot\frac{1}{1-\lambda_2 h_1(M)}, \tag{2.5}$$

Пусть W_i, i 1= 1,2 время ожидание для вызовов i-го класса тогда

$$E[W_2] = \frac{F}{2h_1(M)} \cdot \frac{1}{1-\lambda_2 h_1(M)}$$

где, $$F = \frac{\lambda_1/\mu^{N-M}}{\prod_{j=0}^{N-M-1}(M+j)} \cdot \frac{2N\mu}{(N\mu-\lambda_1)^3} + 2 \cdot \left(\sum_{k=1}^{N-M-1} \frac{\lambda_1/\mu^k}{\prod_{j=0}^{k-1}(M+j)} h_1(M+k)^2 + h_1 M^2 \right).$$

В [46] оптимизационная проблема для ЦОВ решена путем точных расчетов для относительно небольших ЦОВ. Однако, как показано выше, для ЦОВ с двумя классами вызовов решение является громоздким, сложным и малоэффективным. Поэтому наиболее целесообразно применять асимптотические подходы к решению проблемы оптимизации.

Для асимптотических подходов рассмотрим ряд M/M/{Ki} система пронумерованных r = 1,2, ... (далее верхний индекс) который обозначает r-го систему. Например, $E[W1]$ поддерживает среднее время ожидания высоких клиентов приоритета в r-го системе. Предполагаем, что наша система работает согласно правило Квадратному корню, то есть количество операторов N, и интенсивность нагрузки λ растут следующими образом $\sqrt{N}(1-\rho) \to \beta$, **0 < β <** ∞ если r → ∞.

Асимптотическая вероятность задержки или функция вероятность ожидания Халфина-Витта (Halfln-Whitt function), приведенной в [33] равна

$$\lambda(\beta) = \left[1 + \frac{\beta\Phi(\beta)}{\phi(\beta)}\right],$$

где Φ(*)и ϕ(B) - стандартное нормальное распределение и функция плотности распределения вероятности соответственно.

После этого наблюдения можно описать, в таблице 2.2, соотношения между пороговым уровнем и качеством обслуживания для двух классов вызовов, используя M/M/{Ki} модель.

В частности, для вероятности ожидания пороговый уровень 0 приводит к QED режиму, который характеризуется вероятностями ожидания строго между 0 и 1. При увеличении порогового уровня, время ожидания низкоприоритетных вызовов остается неизменным, в то время как время ожидания высокоприоритетных вызовов уменьшается. Если пороговый уровень увеличивать ещё больше, то время ожидания высокоприоритетных вызовов становится равным нулю (при наличии большого количества операторов в ЦОВ), что приводит в режиму функционирования QD

В таблице 2.2 ρ соответствует интенсивности нагрузки для высокоприоритетных вызовов

Т а б л и ц а 2.2. Уровень качества обслуживания для обоих приоритетов

	Пороговый Уровень	$\sim P\{W_1^r>0\}$	$\sim P\{W_2^r>0\}$	$E[W_1^r \mid W_1^r>0]$	$E[W_2^r \mid W_2^r>0]$
A	0	$0<\alpha(\beta)<1$	$0<\alpha(\beta)<1$	$\theta(1/N)$	$\theta(1/\sqrt{N})$
B	B	$\alpha(\beta)\cdot\rho_1^b$	$\alpha(\beta)$	$\theta(1/N)$	$\theta(1/\sqrt{N})$
C	$c.\ln N$	$\alpha(\beta).\rho_1^{c\ln N}$	$\alpha(\beta)$	$\theta(1/N)$	$\theta(1/\sqrt{N})$
D	$d.\sqrt{N}$	$\theta(\alpha(\beta-d)\rho_1^{d\sqrt{N}})$	$\alpha(\beta-d)$	$\theta(1/N)$	$\theta(1/\sqrt{N})$

Из таблицы. 2.2 можно видеть, как пороговый уровень влияет на качество обслуживания двух классов.Выше сказано что β - есть затраты на обслуживания, в данном параграфе будим определить как минимизировать эти траты. Такой подход имеет точки зрения минимизации расходов на персонал в связи с различными факторами, например, класс зависит от оценки среднего времени ожидания, и о вероятности ожидания более чем предварительно заданного времени. Кроме того, этот подход стремится свести к минимуму общую стоимость, которая представляет собой сумму расходов на персонал и расходы, связанны с ожиданием.

Рассмотрим еще раз модель изображенную на рисунке 2.3, с J классами вызовами. Вызовы класса i поступают согласно пуассоновскому закону с интенсивностью λi не зависимо от остальных классов. Длительности обслуживания предполагаются распределенными по экспоненциальному закону, а интенсивность обслуживания ц для всех классов вызовов. Пусть вероятность ожидания для различных классов вызовов обозначается как $P\{Wi>0\}$ для i -го класса, и пусть $0<\alpha i<1$, i=1, ….J. Пусть классы вызовов расположены в возрастающем порядке в соответствии с α_1, $\alpha_2 < \alpha_2 <.... < \alpha$,. Пусть также П будет обозначать набор всех правил маршрутизации, которые не предполагают прерывание обслуживания. При заданном правиле маршрутизации π-П, пусть $P_\pi Wi>0$ будет стационарной вероятностью того, что вызов i-го класса будет ожидать в очереди. Решение проблемы минимизации затрат на операторов заключается в подборе оптимального их количества.

Как альтернатива задаче (2.3) можно рассмотреть проблему повышения прибыли, в которой вызов класса i приносит доход, который уменьшается с увеличением его времени ожидания. Обозначим этот доход как r_i-$c_i E[Wi]$.

Следовательно,интенсивность доходов, приносимых определенным классом вызовов, будет $r_i\lambda_i$-$c_i E[W_i]$. Поскольку величина $r_i\lambda_i$ не зависит от

выбора модели для расчета операторов и управления, итоговая проблема повышения доходов будет эквивалентна (2.3) .

Как упоминалось ранее, оптимальный подход, пытаясь свести к минимуму затрат, заключается в использовании пороговых приоритетных уровней. Существует достаточно доказательств, что этот принцип применяется гораздо шире [20, 25, 22, 23]. Более того, даже если подход оптимально, все равно придется, определить кадрового уровень и оптимального порога. Возможно, использовали работу, сделанную в [26], чтобы решить проблему оптимизации (2.2) прямым перечислением для систем разумно небольших размера. Однако, как показано в случае с двумя классами, это очень усложнено, трудоёмкое и вряд ли обеспечит полезные информации. Поэтому наиболее целесообразно применять асимптотические подходы к решению проблемы оптимизации. Обычно за основу берется предельный случай, когда загруженность ЦОВ высока.

Выше были рассмотрены методы решения задач по расчету количества операторов и методов управления для ряда ЦОВ. В действительности, имеем один ЦОВ с прогнозируемой нагрузкой, заданными пределами для времени ожидания вызовов, рассчитанным штатом операторов и затрат при ожидании. Как применить результаты, полученные выше к такому ЦОВ? Анализ предполагает, что если общая интенсивность вызовов высока, класс вызовов J имеет сопоставимый количественный размер, и затраты при ожидании для этого класса сопоставимы и не превышают затраты на операторов, тогда правило квадратного корня и пороговых условий является асимптотически оптимальным [16,24]. Описанных подходов, являются весьма перспективными и в то же время каждый из них есть свои тонкости и проблемы.

Учитывая сложность общей крупномасштабных систем обслуживания, трудно оценить применимость порогово-приоритетного правила управления для этих систем. Подход, который может привести к простым правилам управления, является то, что взгляд на простые схемы маршрутизации (которые могут оптимизировать другие критерии эффективности - другие, чем минимизация затрат).

Другой важный практический вопрос, который может быть решен при использовании данной модели - конфигурация ЦОВ. Например, может рассматриваться две возможных конфигурации для мультисервисного ЦОВ Первая 1-модель, в которой каждый класс вызовов обслуживается своей группой операторов. Вторая У-модель (рисунок 2.3), в котором все операторы обладают необходимыми навыками, чтобы обслуживать все классы вызовов. Очевидно, что при одинаковом уровне обслуживания в У-модели потребуется меньшее количество операторов, чем в 1-модели. Но насколько? Также можно определить максимальные затраты на дополнительное обучение операторов, которые будет оправдывать переход с 1-модели на У-модель.

Дисциплина обслуживания - это способ определения того, какое требование в очереди должно обслуживаться следующим. Решение может основываться на одной из приведенных ниже характеристик или на их совокупности:

– мера, определяемая относительным временем поступления рассматриваемого требования в очередь;

– мера требуемого или полученного до сих пор времени обслуживания;

– функция, определяющая принадлежность требования к той или иной группе.

Приоритет запроса - характеристика, определяющая место запроса в очереди на его обслуживание. Приоритет назначается в соответствии с характером задачи, решаемой по запросу или по роли источника запроса.

При выборе дисциплины обслуживания следует учитывать следующие требования:

– запросы высшего приоритета должны обслуживаться в кратчайшие сроки;

– запросы низшего приоритета должны обслуживаться в приемлемые для абонента сроки;

– должна быть обеспечена относительная простота реализации выбранной дисциплины обслуживания;

– должна быть обеспечена максимум полезной работы (т.е. обслуживания запросов), минимум потерь на переключение;

– должно быть обеспечено уменьшение среднего времени реакции на запрос и среднего числа запросов в очереди.

Эти требования взаимоисключающие. Следовательно, нужно искать компромисс, оптимум при определенных условиях.

Различают две группы дисциплин обслуживания запросов :

а) без приоритета;

б) с учетом приоритета.

Абсолютный приоритет обеспечивает прерывание процесса обработки текущего требования в случае поступления на вход устройства заявки с более высоким приоритетом. Требование, находившееся на обработке, устанавливается в очередь к устройству и занимает в ней место согласно его приоритету.

При относительном приоритете вновь поступившие требования всегда становятся в очередь к устройству согласно своим приоритетам. В отличие от абсолютного приоритета при использовании данного механизма новая заявка с более высоким приоритетом не может прервать процесс обработки текущего требования, даже если последнее менее приоритетное. Прибывшая заявка устанавливается в очередь, а по окончанию обработки на обслуживание
принимается требование, обладающее наивысшим приоритетом.

При использовании смешанного дисциплина обслуживания в момент поступления более приоритетного требования на вход занятого устройства производится выбор абсолютного или относительного способа учета приоритетов. Если планируемое время обрабатываемой заявки меньше значения, определяемого параметром, то обслуживание текущей заявки продолжается, в противном случае - прерывается. На рисунке 2.4 представлена иллюстрация разных дисциплин обслуживания.

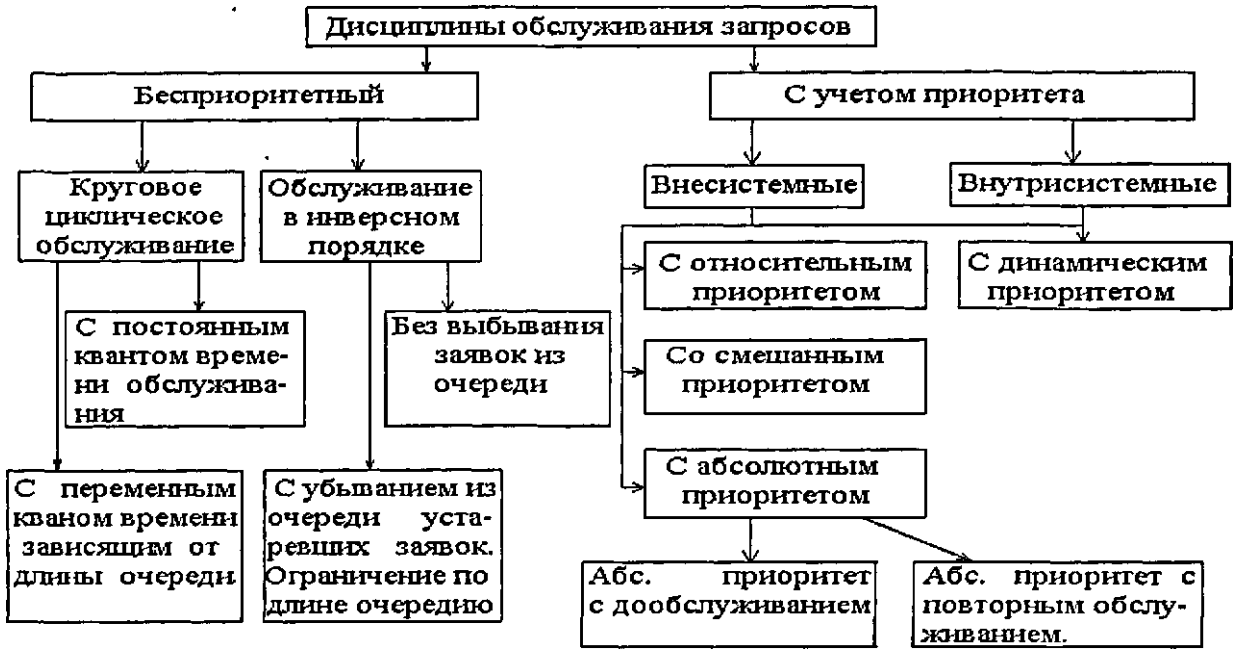


Рисунок 2.4 - Классификация дисциплин обслуживания запросов

2.5 Общая модель МКЦ с относительными приоритетами

Пусть имеется p потоков запросов на предоставление информационных услуг, которые поставлены в соответствие p приоритетов. Поступающая нагрузка создается p разнотипными потоками с показательными распределениями интервалов времени между запросами и параметрами = 1. Для речевых запросов, приходящих от телефонных сетей, сетей VoIP и СПС такое предположение является общепринятым, исходя из основополагающих работ Эрланга и Энгсета [4, 20, 21]. Для запросов других типов такое предположение так же можно считать приемлемым, основываясь на результатах раздела 2.2 (таблица. 2.2).

Классификация потоков запросов строится по типам поступающей информации и допустимому времени на её обработку. Нагрузку на операторскую подсистему и контакт-центр в целом создают запросы от абонентов телефонных сетей, СПС, пользователей социальные сети, сетей VoIP и интернет. Различаются запросы и по видам представления - речевые, текстовые, требующие непрерывной обработки и допускающие отложенное

обслуживание. Система представлена совокупностью отдельных модулей, каждый из которых в один момент времени может обслуживать лишь один запрос. В зависимости от типа поступившего запроса и с учетом приоритетности обслуживания, он попадает в ту или иную очередь накопителя, а затем поступают на обслуживание. Модель изучаемой в данной главе представлена на рисунке 2.5.

Здесь λ_i - интенсивности поступления на подсистему запросов различных типов, где i= (1…. p) — индекс типа запроса.

Накопитель заявок реализует приоритетную дисциплину обслуживания, в зависимости от неё в состав накопителя входит определенное число очередей заявок различных приоритетов p. В общем случае каждому типу запросов соответствует свой уровень приоритета и выполняется условие p < i

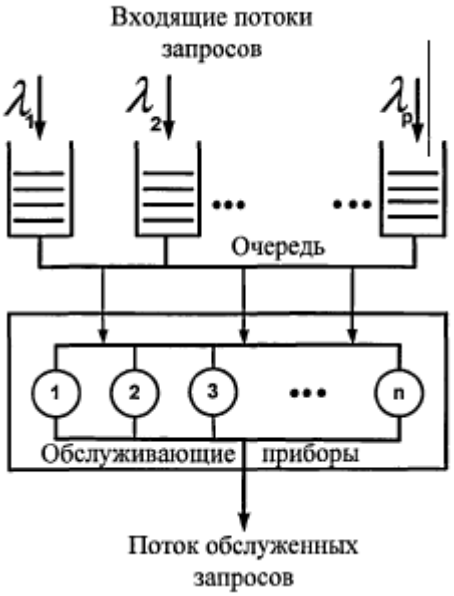


Рисунок 2.5 - Функциональная модель операторской подсистемы МКЦ с приоритетным обслуживанием запросов.

Заявки, поступающие в систему, обслуживаются одной группой операторов. Если в группе имеется несколько свободных операторов, то заявка поступает на обслуживание к тому оператору, который был свободен дольше других. Если в группе есть хотя бы один свободный оператор, то заявка, требующая обслуживания поступает к этому оператору. Если в момент поступления заявки все операторы заняты, то заявка помещается в соответствующую ее происхождению очередь. Каждой очереди заявок присвоен определенный приоритет в обслуживании. Для упрощения модели будем считать, что производительность всех операторов в группе одинакова. Но для времени обслуживания запросов, поступающих из разных очередей, аналогичное упрощение недопустимо.

Процесс обслуживания заявок отдельным рабочим местом оператора (РМО) контакт-центра предполагается описывать СМО с относительным приоритетом; уже начатая процедура обслуживания доводится до конца, даже если во время ее реализации в систему поступает требование с более высоким приоритетом. Запросы, имеющие одинаковый приоритет, обслуживаются по принципу «первым пришел - первый обслужен».

Необходимо отметить, что характеристика дисциплины очереди - обслуживание в соответствии с приоритетом без прерывания играет решающее значения по сравнению с характеристикой ограничения поступления заявок в очередь (закрытие очереди при превышении граничного времени ожидания обслуживания). Сказанное справедливо при выполнении следующих условий:

- число обслуживающих приборов (в данном случае количество рабочих мест операторов) достаточно для того, чтобы время ожидания заявки в очереди не превысило граничного времени ожидания;

- приоритеты обслуживания очередей установлены таким образом, что первыми обслуживаются заявки из очередей с меньшим граничным временем ожидания. В нашем случае - это очередь, телефонных вызовов. Из приведенной моделирования операторской подсистемы видно, что

для рассмотрения её в целом необходимо определить показатели эффективности работы системы которые будут получены в результате анализа. Эта задача решается в следующих разделах.

2.6. Количественная оценка характеристик приоритетных моделей обслуживания мультисервисных контакт-центров

Предположим, что требования, сформулированные в предыдущих разделах, соблюдены. В этом случае математической моделью исследуемой системы, может быть СМО и различными приоритетами обслуживания заявок. Используя стандарт обозначений Кендалла-Ли, рассмотрим СМО вида

$$(M/M/n): NPRP/\infty/\infty),$$

где М - пуассоновское распределение моментов поступления заявок на обслуживание и экспоненциальное распределение продолжительностей обслуживания заявок;

n - число рабочих мест операторов в системе;

NPRP - дисциплина очереди, не допускающая прерывания обслуживания уже принятой к исполнению заявки;

∞ - означает, что максимальное число допускаемых в систему заявок и емкость источника, генерирующего заявки на обслуживание, не ограничены.

Предположим, что поступающие заявки принадлежать одному из p различных приоритетных классов, обозначаемых через p (p = 1, 2,...., P). Будем считать, что чем меньше индекс класса тем, выше приоритет этого

класса. Отметим, что каждый типов запросов различные классы имеет свое время обслуживания. Исследуемая модель требуется общего времени обслуживания, для этого, нужно брать среднее полное время обслуживания всех типы запросы различные классы.

В рамках данной работы искомыми характеристиками качества предоставления информационных услуг для операторской подсистемы принято среднее время ожидания и среднее время пребывания запроса к в подсистеме МКЦ.

Среднее время пребывания в системе запроса приоритета к обозначается через $T_к$. Вводятся обозначения: $У/_к$ - среднее время ожидания начала обслуживания заявки на предоставление информационной услуги приоритета к, £ - среднее полное время обслуживания поступившего в МКЦ всех типы запросы различные классы, т.е. время о начала предоставления информационной услуги до завершения.

Очевидно,

$$T_к = УГ_к + 5 . \qquad (2.4)$$

Рассмотрим время ожидание для требования к-го приоритета Щ. Время ожидания для заявки разлагается на три составляющие [12]

$$W_k = V + \sum_{i=1}^{k} \frac{S}{n} \cdot L_i + W_k \cdot \sum_{i=1}^{k-1} \frac{S}{n} \lambda_i \equiv V + W_N + W_M \qquad (2.5)$$

где V - время связанное с тем, что в момент поступления данной заявки другая заявка находится в обслуживающем приборе;

а) время, обусловленное заявками, находящимися в очереди в момент поступления данного требования данной заявки и на наконец;

б) время, связанное с данной заявки, поступающими позже данной заявки.

Исследуем систему в состоянии вновь поступившей заявки из приоритетного класса к. Будем называть эту заявку меченой. Первая составляющая времени ожидания V для меченой заявки связана с другой заявкой, которую она застанет обслуживанием приборе; эта составляющая равна остаточному времени обслуживания другой заявки другого требования, а распределение времени обслуживания зависит от приоритета другой заявки . Вторая составляющая времени ожидания , это именно задержку, связанную с тем, что перед меченой заявкой обслуживаются другие заявки, которые меченая заявка застало в очереди [13].

Аналогично можно определить третью составляющую среднего времени ожидания Wm (задержку меченой заявки меченого требования за счет заявок, поступающих после неё).

Введем несколько обозначений: интервалы времени между поступлениями заказов с приоритетом распределены в соответствии с законом Пуассона и характеризуются средней частотой поступления, равной λ_i 0=1, 2, . . ., p); средние значения длительности обслуживания все заявки определяет как μ

$S=\mu^{-1}$полное среднее время обслуживания; (2.6)

$A=\lambda \cdot S$ полная поступившая нагрузка **(2.7)**

Для предложенной модели существует еще более простая формулировка определения среднего времени пребывания заявки в очереди соответствующего k-го приоритета.

2.7. Расчеты и приоритетная стратегия обслуживания запросов на информационные услуги

Для расчетов по предложенному методу можно использоваться пакетом прикладных программ для математических расчетов МаШСАБ. Приведем пример использования подобной методики для исследования характеристик отдельного элемента обслуживания всей операторской подсистемы. Принятие приоритетной стратегии во многом зависит от каждого конкретного случая внедрения центра информационных услуг, можно выделить следующие факторы распределения поступающих на подсистему запросов по приоритетам:

а) по важности, в зависимости от идентификатора пользователя;

б) по важности, в зависимости от идентификатора службы;

в) по срочности начала и скорости обслуживания в зависимости от типа запроса.

Очевидно, что для практической реализации важны все факторы, на их примере и рассматривается операторская подсистема современного контакт-центра, осущесвляюшая взаимодействие с множеством инфокоммуникационных сетей. На рисунке 2.6 приводится пример распределения поступающих на изучаемую подсистему заявок по приоритетам.

Такой выбор может быть обоснован тем, что в случае поступления в МКЦ запросов первых двух приоритетов они, попадая в очередь, занимают жестко ограниченный ресурс - каналы телефонной сети с коммутацией каналов, и, как следствие, требуют скорейшего обслуживания. Кроме того, как правило, при речевом запросе пользователь готов ждать меньше, чем при запросе текстовом. Поэтому же речевые вызовы от СПС имеют более высокий приоритет, требуют скорейшего обслуживания.

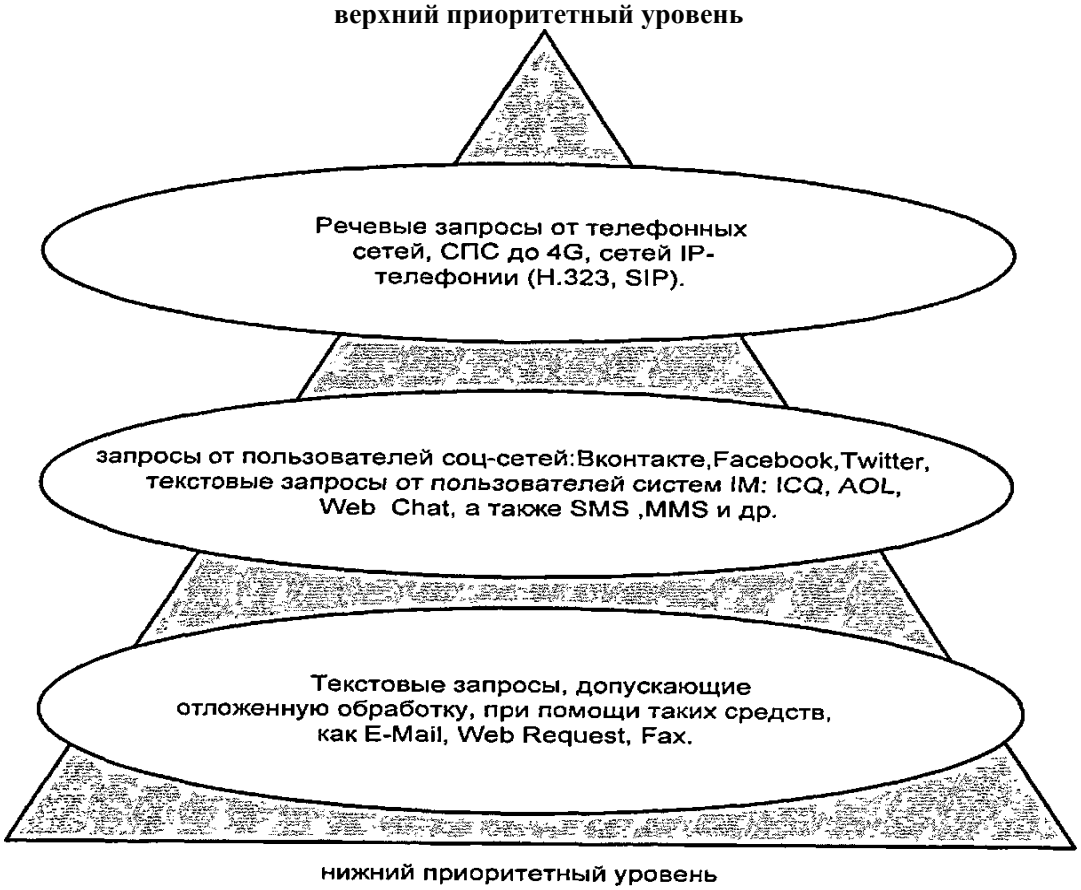


Рисунок 2.6 - Схема распределения заявок по приоритетам

Таким образом, предполагается приоритетная стратегия.

Первый уровень приоритетов - речевые запросы и он включает:

- вызовы от пользователей сетей подвижной связи;

-вызовы от пользователей, фиксированных телефонных сетей;

-запросы от пользователей сетей IP- телефонии.

Второй уровень приоритетов соответствует:

- запросам пользователей социальных сетей (например, Вконтакте, Facebook, Twitter);

- запросам пользователей систем мгновенного обмена текстовыми сообщениями - IM. В данном случае рассматриваются запросы от систем IM (ICQ, AOL и подобные); запросы через Web Chat, наконец,

- запросы через службу SMS, MMS от пользователей СПС (например, картинки, фотографии, видеоролики, а также писать текстовые сообщения длиной более 1000 символов).

Подробное разделение типов приоритетов второго уровня весьма условно и во многом зависит от конкретной ситуации, в частности уровня обслуживания в СПС, от которого зависит возможность отнести службу SMS к системам IM. Так же условно разделение между службами IM и средствами Web Chat, хотя в отдельных случаях различие между ними может позволить

определить уровень заинтересованности пользователя в информационной услуге и постоянство его обращения к услугам контакт-центра.

Третий уровень приоритетов запросы, позволяющие отложенную обработку. К ним относятся заявки, поступающие по электронной почте или в виде факсимильных сообщений. В качестве примера предполагается, что заявки этих типов всегда обслуживаются в последнюю очередь.

2.8 Сравнение приоритетной и бесприоритетной организации процессов предоставления информационных услуг

На рисуке 2.7 приведены графики зависимости времени ожидания в очереди при разных значениях суммарной нагрузки p запроса разного типа приоритетов.

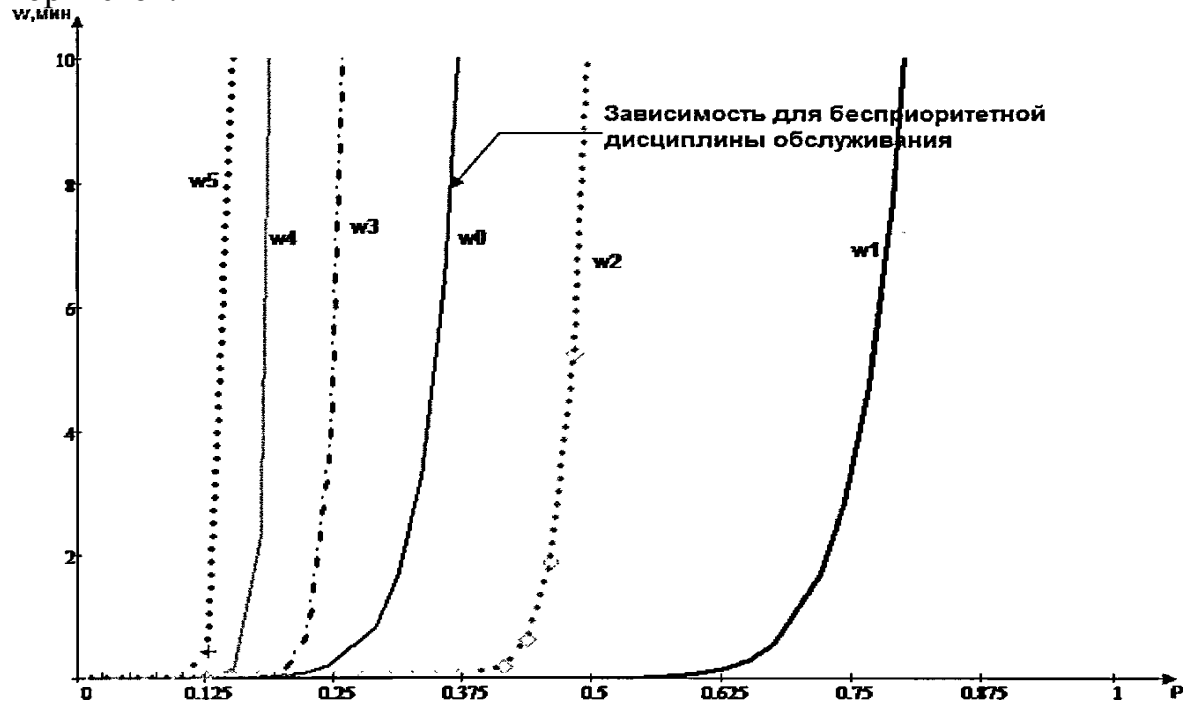


Рисунок 2.7 - Зависимости среднего времени ожидания запросов разного типа в системе от её суммарной загрузки для приоритетной и бесприоритетной дисциплин обслуживания

Полученные для базовой модели математические зависимости характеристик обслуживания заявок от параметров системы при использовании приоритетных дисциплин обслуживания заявок позволяют выполнить достаточно полный анализ свойств исследуемой системы. Так, например, анализ влияния суммарной загрузки системы на характеристики обслуживания заявок показывает, что среднее время ожидания в очереди заявок всех классов растет с увеличением суммарной загрузки $\boldsymbol{p}$, причем более резко в области больших значений загрузки, особенно для заявок низкоприоритетных классов. В области перегрузок, когда $\boldsymbol{p} > 1$, проявляется

свойство защиты от перегрузок высокоприоритетных заявок за счет отказа в обслуживании низкоприоритетным заявкам. При этом время пребывания низкоприоритетных заявок возрастает неограниченно и стремится к бесконечности, в то время как для высокоприоритетных заявок время пребывания имеет конечное значение.

Разработано формализованное описание операторской подсистемы контакт-центра основными элементами.

Исследованы потоки запросов, поступающие в операторскую подсистему МКЦ от различных источников, определены их особенности и особенности их совместной обработки.

3 Исследование МКЦ с отложенным обслуживанием заявок на информационные услуги

3.1 Алгоритм функционирования мультисервисных контакт-центров с отложенным обслуживанием заявок

Вернемся к описанию идей off-line МКЦ в главе 1.4 и рассмотрим основные алгоритмы функционирования операторскую подсистему. Алгоритм функционирования каждой конкретной реализации контакт-центра включает в себя базовую структуру, которая описана ниже. Клиент посылает запрос операции с использованием одного из доступных методов, таких как голосовые сообщения, SMS, MMS, электронная почта, мгновенные сообщения, веб-форму запроса и т.д. Запрос может быть сформулирован как вопрос свободной форме. Поставщик услуг может зарезервировать за собой право не отвечать на некоторые типы вопросов. Это требование переходит к оператору.

Маршрутизация запросов к оператору осуществляется автоматически, с использованием ряда условий, таких как состояние занятости оператора, его опыт и специализация. Обслуживание запроса, как правило, базируется на поисковой системе общего пользования (интернет). Когда ответ готов, оператор сообщает об этом потребителю различными способами, например, при помощи системы интерактивного обмена текстовыми сообщениями web-чат, SMS, MMS, электронной почты, и т.д.

Алгоритм обработки запроса представлен на рисунке 3.1.

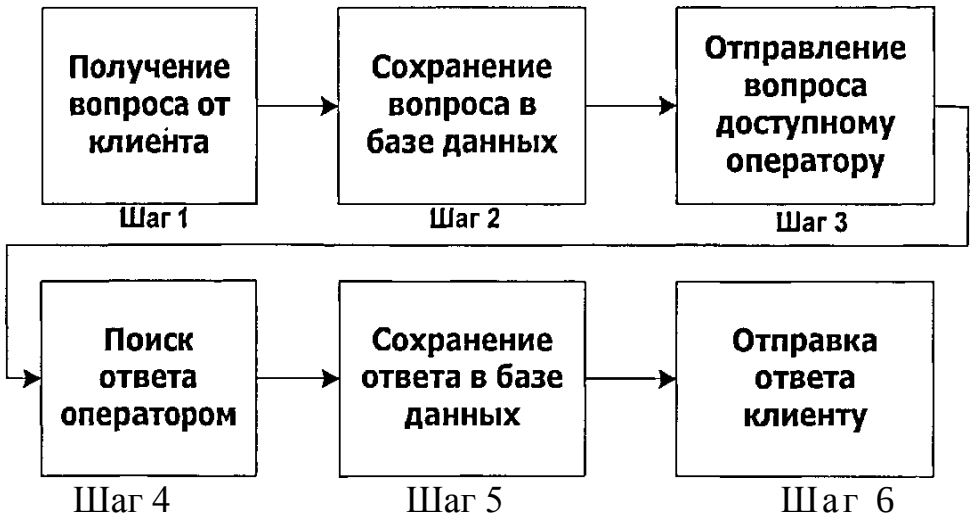


Рисунок 3.1 - Алгоритм обработки запроса для системы с отложенным обслуживанием

Он характеризуется следующими основными шагами:

- система получает свободной форме вопрос от клиента по SMS;
- вопрос в настоящее время сохраняется на базе данных;

- система отправляет вопрос доступному оператору, равномерно загружая операторов системы;

- оператор ищет ответ на вопрос с помощью поисковых систем общего пользования, доступных в сети Интернет;

- оператор формирует ответ и посылает его для сохранения в вазу данных;

- посылка ответа клиенту.

Если в разделе 2.2, в качестве одного из основных положение исследовании контакт-центры, допускали что все поступающие заявки, обслуживаются по показательному закону то здесь нельзя такие допущении, поскольку в этом частном случай работы оператора ведется эксклюзивном с интернетом. Рассмотрим характеристики этот трафик.

3.2 Анализ Интернет трафика

Трафик с коммутацией пакетов, таких как Интернет, было показано, что бы дать понятия как их характеристики, сильно отличаются от традиционного трафика телефонной сети[21].В частности, это пульсирующий на различных масштабах времени, показывает, дальний, а не ближний зависимость в автокорреляции, и статистически автомодельным, т. е. фрактала. Это влияет на характеристики сети в неблагоприятных пути, и усложняться проектирования сетей. В старые добрые времена проектирование сети, жизнь была проще. Существовал только один вид движения никакого значения, и это был голос. Он был (и есть) хорошо известными характеристиками, а именно Пуассоновский входящий потока и экспоненциального распределения длины вызова. Существовал не нужно беспокоиться о таких вещах, как сеть слоев, так как они не существуют. Это было легко измерить критических значений важных параметров.

Теория массового обслуживания разрешается дизайн голосовых сетей для удовлетворения любой желаемой характеристики. Одной из проблем в расчете вероятностно-временных характеристик сетей, возникшей в последние годы, является необходимость адекватного учета реального характера трафика в сетях. Долгое время считалось, что природа сетевого трафика соответствует Пуассоновскому процессу. К сожалению, эти времена прошли навсегда. Со временем количество исследований и измерений характеристик сетевого потока возрастало.

В результате было замечено, что не всегда поток пакетов в локальной или глобальной сети можно моделировать с использованием Пуассоновского процесса. С пакетной сети, многослойные протоколов здесь остаться. Это означает, что многие более инварианты, как один или несколько, как правило, связаны друг с сетевого уровня, гораздо более сложная статистика трафика, и, соответственно, гораздо труднее анализа и моделирования сетевого трафика. Последние исследования различных типов сетевого трафика убедительно доказывают, что сетевой трафик является самоподобным или фрактальным по

своей природе, т.е. в нем присутствуют так называемые вспышки или пачки пакетов, наблюдаемые в различных временных интервалах (от миллисекунд до минут или даже часов). Из этого следует, что широко используемые в настоящее время методы моделирования и расчета сетевых систем, основанные на традиционных предположениях, не дают адекватной картины происходящего в сети. Стандартное предположение о том, что потоки информации в сетях являются стандартными пуассоновскими, оказывается неверным, поскольку эти потоки уже не являются суперпозицией большого числа независимых стационарных ординарных потоков равномерно малой интенсивности. Поэтому ведется работа по созданию адекватных моделей реального трафика, который имеет взрывной характер и корреляцию длин интервалов между моментами поступления запросов в систему, на основе самоподобных процессов.коммутации каналов.

Характеристики трафика в данных сетях хорошо изучены, а также разработаны строгие методики расчетов. В основу компьютерных сетей, как правило, был положен принцип коммутации пакетов, а методики расчетов, возможно, вследствие некоторого отставания теоретической базы от бурно развивающихся технологий остались практически теми же, что и привело к возникновению "проблемы самоподобия".

Наглядной особенностью самоподобного трафика в контексте производительности сетей является «устойчивость кластеризации». В пуассоновском трафике кластеризация наблюдается в краткосрочном масштабе времени, а в долгосрочном сглаживается. Так, если спроектировать систему из серверов и очередей в расчете на долгосрочное сглаживание, то будет достаточно буферов умеренного размера, т.к. очередь может образоваться в краткосрочной перспективе, но за долгий период времени буферы очистятся. На практике оказывается, что традиционный анализ очередей, в основе которого лежит предположение о пуассоновском потоке, не всегда может точно предсказать производительность системы в условиях самоподобного трафика.

Параметр H, называемый коэффициент Хэрста (Hurst parameter), или параметром самоподобия, имеет принципиальное значение в теории самоподобных процессов. Он является индикатором степени самоподобия процесса, а также свидетельствует о наличии у него таких свойств как персистентность/антиперсистентность и продолжительная память. Используя значение показателя Хэрста H, выделяют три типа случайных процессов:

В случае 0.5<H<1 говорят о персистентном (поддерживающемся) поведении процесса, либо о том, что процесс обладает длительной памятью. Другими словами, если в течение некоторого времени в рошлом наблюдались положительные приращения процесса, то есть происходило увеличение, то и впредь в среднем будет происходить увеличение. Иначе говоря, вероятность того, что процесс на 1+1 шаге отклоняется от среднего в том же направлении, что и на \ шаге настолько велика, насколько параметр H близок к 1. Таким

образом, персистентные стохастические процессы обнаруживают четко выраженные тенденции изменения при относительно малом "шуме".

В случае 0<H<0.5 - случайным процесс является антиперсистентным, или эргодическим, рядом, который не обладает самоподобием. Здесь высокие значения процесса следуют за низкие, и наоборот. Другими словами, вероятность того, что на 1+1 шаге процесс отклоняется от среднего в противоположном направлении (по отношению к отклонению на \ шаге) настолько велика, насколько параметр H близок к 0.

При H=0.5 отклонения процесса от среднего являются действительно случайными и не зависят от предыдущих значений, что соответствует случаю броуновского движения.

Основное свойство случайной величины, распределенной в соответствии с РТХ, состоит в том, что она проявляет высокую изменчивость. Иными словами, выборка из РТХ представляет собой большей частью относительно небольшие значения, однако также содержит и достаточное количество очень больших значений. Для распределения Парето параметр ***к*** определяет минимальное значение, которое может быть принято случайной переменной. Параметр ***а*** определяет среднее значение и дисперсию случайной переменной. Если **а** <2, тогда распределение обладает бесконечной дисперсией, если ***а*** < 1, то распределение будет обладать бесконечным средним значением и дисперсией. На рисунке. 3.3 представлена форма распределения Парето для параметров **к=2** и д=1.5 со средним значением 6 и бесконечной дисперсией.

Согласно [20], если некоторая величина η имеет нормальное (0,1) распределение, то случайная величина x: = exp(ση+ *n)* имеет логнормальное распределение с параметрами (m,σ^2). Для логнормального распределения **плотность** распределения вероятности определяется, как

$$M(x(t)) = exp(\frac{1}{2}\sigma^2 + m), \quad (3.7)$$

$$MX^{(z)} = \exp(\frac{1}{2} \cdot z^2 \cdot \sigma^2 + zm), \quad (3.8)$$

$$D(x(t)) = e^{2+2m}(e^{\sigma 2} - 1). \quad (3.9)$$

При H=0.5 отклонения процесса от среднего являются действительно случайными и не зависят от предыдущих значений, что соответствует случаю броуновского движения.

На рисунке 3.4 представлена форма логнормального распределения для параметров i=1.3 и a =1 со средним значением 6.05 и дисперсией 62.89.

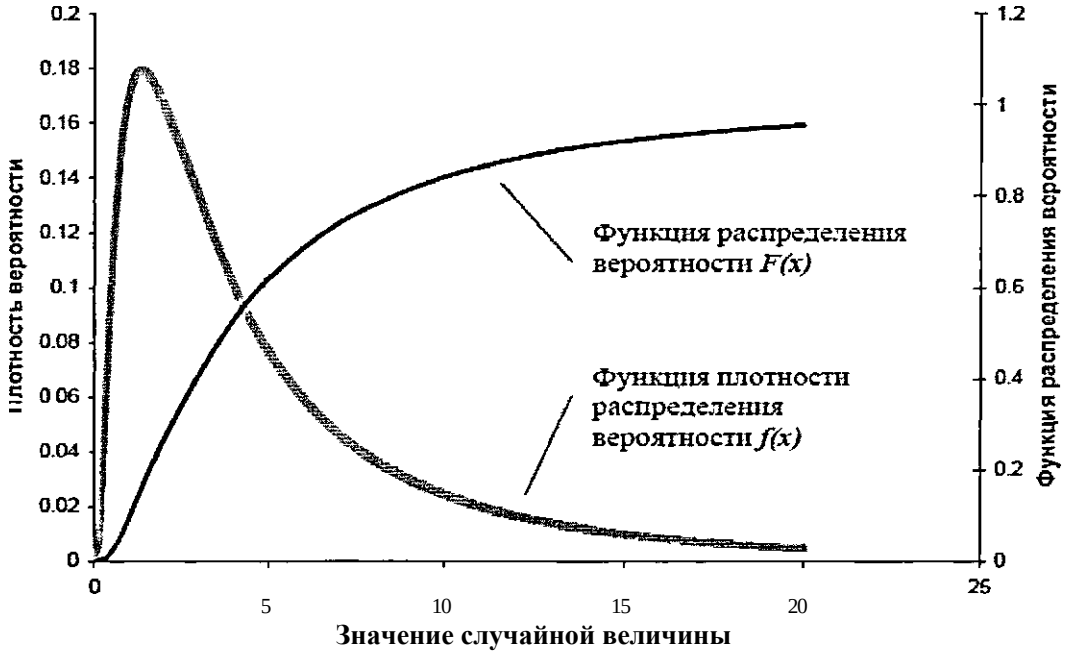


Рисунок 3.4 - Логнормальное распределение.

Случайная величина X имеет распределение Вейбулла - Гнеденко с двумя параметрами *a* и *Ь,* если ее функция распределения и функция плотности вероятностей имеют соответственно вид

$$F(x;a,b)=(\frac{b}{a})\cdot(\frac{x}{a})^{b-1}\cdot \exp(-\frac{x}{a}), \quad (3.10)$$

где a >0 - параметр масштаба, b >0- параметр формы

$$F(x;a,-b)=1-\exp(-\frac{x}{a})^{b},\ x>0, \quad (3.11)$$

С тремя параметрами ее функция распределения и функция плотности вероятностей имеют соответственно вид

$$F(x;a,b,c)=\begin{cases} \frac{b}{a}(x-c)^{b-1}exp\left|-(\frac{x-c}{a})^{b}\right|; x\geq c, \\ 0; x<c \end{cases} \quad (3.12)$$

$$F(x;a,b,c)=\begin{cases} 1-exp\left|-(\frac{x-c}{a})^{b}\right|, \\ 0; x>c \end{cases} \quad (3.13)$$

где *a* >0 - параметр масштаба;

b>0- параметр формы;

c - параметр сдвига.

Экспоненциальное распределение - весьма частный случай распределения Вейбулла - Гнеденко, соответствующий значению параметра формы b = 1.

Нарисунке 3.6 представлена форма распределения Вейбулла-Гнеденко, для параметров а = 2 и Ь=2 ,с=1 со средним значением 2.772 и дисперсией 0.858.

Современные исследования показывают, что самоподобность в идеализированном окружении (т.е. с неограниченными ресурсами и независимыми источниками трафика) может возникать в результате объединения множества отдельных, хотя и сильно изменчивых ON/OFF источников (т.е. ON и OFF-периоды имеют распределения с тяжелыми хвостами (PTX) и бесконечные дисперсии, например распределения Парето).

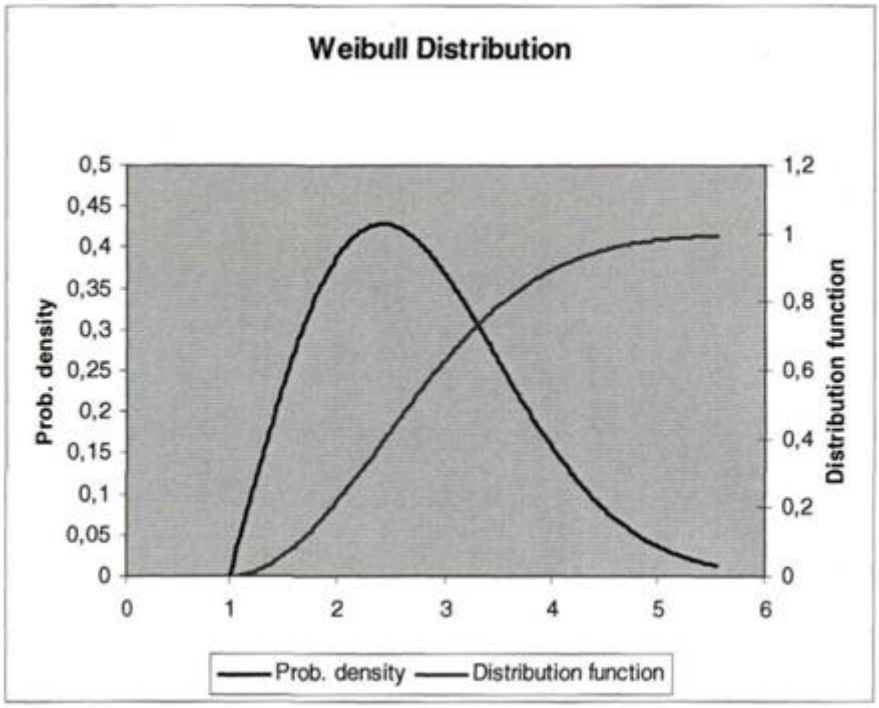


Рисунок 3.6 - Распределение Вейбулла – Гнеденко

Другими словами, наложение множества ON/OFF-источников, проявляющих синдром бесконечной дисперсии («эффект Ноя»), в результате даёт самоподобный объединенный сетевой трафик, стремящийся к фрактальному броуновскому движению («эффект Джозефа»). Кроме того, исследование различных трафиковых источников показывает, что высокоизменчивое поведение ON/OFF - это свойство, присущее архитектуре Клиент/сервер.

Эти результаты, несмотря на их ценность, основываются на предположениях, которые не являются реалистичными в условиях сетевого окружения. Несмотря на приведённые выше рассуждения, реальные клиент/серверные сетевые окружения предполагают ограниченность ресурсов. Это означает, что из-за борьбы за ограниченные ресурсы могут возникать нелинейные процессы, и как следствие возникает связь между источниками трафика. Кроме того, различные (основанные

на обратной связи) механизмы управления, например алгоритм планировщика OS (Operating System), TCP,Ethernet, также могут давать дополнительную нелинейность в случае перегрузки.

Сложность понимания лежащих в основе принципов, которые могут привести к самоподобности в сетевом трафике, в основном определяется тем, что не существует одного причинного фактора, вызывающего самоподобность. Различные корреляции, существующие в самоподобном сетевом трафике, которые воздействуют на различных временных масштабах, могут возникать по различным причинам, проявляя себя в характеристиках относительно конкретного временного масштаба. Это может быть, например, структура информации и поиска (приложений, диска и программы в памяти), «время обдумывания» пользователя и преимущество передач файлов (сеанс/активность), эффекты кэширования, TCP, Ethernet и различные ATM механизмы управления (управление доступом, управление перегрузками и т.д.). Кроме того, трудности при понимании взаимодействий между различными «отдельными» корреляциями, создаваемыми в сетевом трафике перечисленными факторами, ещё более усложняют проблему.

Перечислены некоторые из основных факторов, которые могут продуцировать в сетевом трафике долговременные зависимости (ДВЗ) различных видов:

1. поведение пользователя;
2. генерация, структура и поиск данных;
3. объединение, трафика;
4. средства управления сетью;
5. механизмы управления, основанные на обратной связи;
6. развитие сети.

Условно все виды услуг, предоставляемые мультисервисной сетью связи, можно разделить на три категории, для которых моделирование процессов будет отличаться друг относительно друга:

а) передача данных;

б) передача речи;

в) видеоконференцсвязь.

Для всех категорий трафик может быть описан самоподобной моделью [22], в качестве которой выбирается модель (ON и OFF -источников с PTX длительностей ON и OFF-периодов.

Количество источников нагрузки для сетей связи довольно велико, поэтому модели должны учитывать наложение большого числа 0К/01Т-источников. Поэтому, согласно предельной теореме для самоподобного трафика [12], все три типа предоставляемых услуг можно представить моделями фрактального броуновского движения (ФБД). Каждой из них будут соответствовать свои параметры и приоритет.

Таким образом, самоподобный трафик (для процесса приращений) моделируется просто как модель суммы фрактального гауссовского шума (ФГШ) и среднего значения с заданной дисперсией и показателем Хэрста.

ФБД/ФГШ модели нашли широкое применение в сетевом проектировании, так как их гауссовость строгое масштабирование позволяют проводить аналитические исследования характера построения очередей.

В [23] для ФБД доказывается, что показатель Хэрста *H* имеет большое значение для оценки длины очереди. Он показывает, неприменимость традиционных моделей трафика, когда реальный трафик является самоподобным. Это, в частности, может служить доказательством широко распространённой точки зрения, что для пакетного трафика без установления соединения коэффициент использования *p* не может быть существенно улучшен за счёт увеличения размера буфера. Кроме того, делается вывод, что каналы передачи с более высокой пропускной способностью могут быть задействованы с большим коэффициентом использования без увеличения размера буфера. Интуитивное объяснение этому кроется в повышенной эффективности при мультиплексировании.

Несмотря на то, что эти результаты справедливы для частного случая, они позволяют получить представление о влиянии самоподобности и ДВЗ на характеристики СМО и в частности 1Р-контакт-центрах. Сложность аналитического описания процессов в СМО с самоподобной входной нагрузкой и несколькими приоритетами, и как следствие недостаточность проработки данных вопросов, вызывает необходимость проведения имитационного моделирования (глава 4). Это позволит исследовать процессы, протекающие в сетях связи и выявить закономерности их протекания.

3.3 Исследование ВВХ МКЦ с отложенным обслуживанием заявок

Действия оператора, функционирующей в режиме off-line, - это, прежде всего, поиск ответов на запросы с помощью интернета. Следовательно, Web-подсистема КЦ включает интерфейс только одного типа запросов.

Данным типом запросов является посылаемый браузером в соответствующем методе протокола HTTP идентификатор (URI — Uniform Resource Identifier), определяющий нахождение определенного документа в сети в целом и его расположение непосредственно на сервере Web. Под документом будем понимать любой файл или динамически генерируемую информацию, которые могут быть запрошены браузером в процессе предоставления пользователю услуги.

Далее выделим особенности процессов поступления и обработки потоков запросов для операторской подсистемы исследуемого off-line МКЦ. Достаточно адекватно воспроизводит модель off-line МКЦ система телетрафика без потерь вида M/G/n/oo. Исходя из параграфа 2.2 главы 2 и согласно экспериментальным исследованиям, приведенным в работах Dr Thomas B.Fowler в [21, 22, 23], оказывается, что поступающие потоки в подсистемы off-line КЦ подчиняются показательному распределению, что отражает символ "M" в приведенном описании. Нужно еще отметить, что в ряд экспериментальных работ содержит утверждения о показательном

характере распределения интервалов времени между поступающими запросами (Арлитт [53-55]).

Время обслуживания представляет собой произвольное распределение (G), потому что вопросы сформулированы в свободной форме. Символ "n" указывает количество рабочих мест операторов в системе. Последний символ описывает бесконечное число мест для ожидания, что объясняется большими размерами входных буферов современных систем. Функциональная модель offline МКЦ дана на рисунке 3.7.

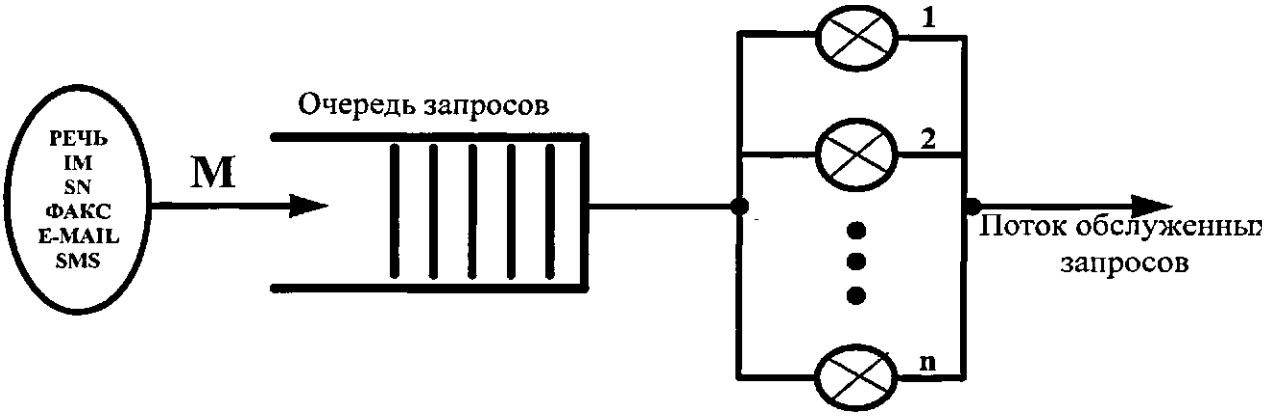


Рисунок 3.7 - Функциональная модель МКЦ с отложенным обслуживанием

Отдельным вопросом стоит проблема моделирования процессов обслуживания Web подсистемой запросов от терминального ПО пользователей. Значительный объем исследований в этом направлении был проведен в работах [16, 17], а также [24, 25] и некоторых других. Исследователи показали, что время передачи запрашиваемых документов с сервера Web может моделироваться распределениями с тяжелыми хвостами или близкими им распределениями.

Зависимость процесса обслуживания от содержания конкретного сервера выражается в некотором несоответствии экспериментальных данных многих систем, проанализированных, например, в работах М. Кровелла и А. Доуни. Их эксперименты указывают на соответствие процессов обслуживания запросов (распределений времен обслуживания запросов) распределению Парето (P) либо логнормальному распределению (LN). Не смотря на эти результаты, можно сказать, что в нашей работе web подсистема МКЦ нас особо не интересуют, в отличие от операторской подсистемы.

Важной особенностью процесса обслуживания запросов off-line МКЦ является их сильная зависимость от типов вопросов, которые клиенты будут спросить, поскольку длительность формирования ответа зависит от типа запроса. На основании результатов исследования, полученных в [66], можно предположить, что в случае типичных вопросов по некоторым доступным темам, например, в случае запроса информации относительно одного и того же продукта, имеет место марковское распределение время обслуживания. В

случае же простых однородных вопросов имеет место детермированное распределение. В случае поступления более сложных и более разнообразных вопросов в свободной форме, возможно рассмотрение и более сложных распределений. Ссылаясь на [27], можно сказать, что при поступления вопросов в свободной форме распределение времени обслуживания является одним из распределений с «тяжелым хвостом» (логнормальное распределение, распределение Парето и Вейбулла) или близко к некоторому специальному полупараметрическому распределению.

Основываясь, в том же работы [27] и исходя из статических данных, можно сделать вывод о том, что в случае поисковых запросов для различных продуктов и услуг, таких как: адрес информацию о поставщиках и хранить местоположение, цена продукта и спецификации, сервисные характеристики и т.д, можно считать, что более подходящие распределение является логнормальным.

Кроме того, некоторые исследователи предлагают использовать уточненные распределения, например, ограниченное log нормальное (BLN) поскольку поиск не может продолжаться слишком долго, такое искусственное ограничение позволяет в некоторых случаях уменьшить погрешность при моделировании, приблизив моделирующий процесс к реальном, и др, необходимые в случаях сложных реальных процессов. Это приводит к необходимости применения имитационного моделирования, позволяющего оперативно строить нужные модели и обходить технические сложности аппарата аналитического моделирования.

Формально, моделирование сервера Web указанной СМО, с физической точки зрения на происходящие в моделируемом объекте процессы, является недостаточно точным. Подобные СМО подразумевают последовательную обработку поступающих запросов по принципу FIFO. На практике же реализации того или иного сервера Web осуществляют попеременную обработку запросов. Наглядно это может быть описано следующим образом: сначала происходит частичная обработка первой заявки, потом второй, затем сервер возвращается к обработке первой и т.д. Такой подход позволяет сразу нескольким пользователям постепенно получать обслуживание своих запросов, что может быть приемлемо, например, при просмотре больших текстовых файлов.

Однако, основная искомая характеристика, среднее время пребывания запроса в реальной системе, может считаться равной таковой в моделируемой системе с дисциплиной обслуживания FIFO. Для этого надо принять время переключения между обслуживаемыми запросам незначительным, по сравнению с общим временем пребывания в системе.

Таким образом, применение для моделирования систем массового обслуживания вида M/G/n/oo можно считать оправданным.

Приведенные выше данной главы исходные данные позволяют детально изложить и апробировать методы исследования ВВХ операторской подсистемы off-line МКЦ. Начальной задачей данного параграфа является

точное определение искомых характеристик и обоснованных способов их нахождения.

Разработчики контакт-центров всегда интересуются вовремя и вероятностные характеристики. Эти характеристики можно определить множество необходимых вещей, таких как: время ожидания в очереди, общее время обработки, время пребывания, средняя длина очереди, а некоторые вероятностные характеристики, как назначается время превышение вероятности (вероятность избытка времени). В наше случай важно характеристик является время ожидания в очереди и время пребывания в системы. Таким образом, основная задача текущего раздела - исследование ВВХ МКЦ с отложенным обслуживанием, разбивается на несколько частей, средних:

• создание аналитической модели исследуемой подсистемы, включающей обобщенную модель поступающего потока запросов и учитывающей самоподобным свойства времени обслуживания (Модели СМО М/ЦЧ/п,);

• создание имитационной модели операторской подсистемы МКЦ с отложенным обслуживанием, включающей обобщенную модель поступающего трафика и учитывающей самоподобным свойства времени обслуживания подсистемы модели СМО М/М.

Для получения приблизительных аналитических выражений времени ожидания запроса в очереди для моделей СМО вида, M/LN/n можно воспользоваться рядом существующих аппроксимационных формул для времени ожидания в очереди модели СМО M/G/n.

Закон Кингмана для высоких уровней нагрузки при малом числе операторов утверждает[68], что задержка в очереди может аппроксимироваться экспоненциальным распределением. При этом среднее значение ожидания в очереди равно

$$W=W_{M/M/n}\cdot\left(\frac{1+C_2^v}{2}\right), \qquad (3.14)$$

где M/M/n- среднее время ожидания обслуживания запроса в очереди системы СМО типа M/M/n.

Аппроксимация Кингмана для высоко загруженности линий предполагает, что большая часть вызовов будет поставлена в очередь Исходя из формулы (1.5) параграфа 1.2 .можно писать что

$$\boldsymbol{W=E_2(A)\cdot\frac{1}{n}\cdot\frac{1}{\mu}\cdot\left(\frac{1}{1-\rho}\right)\cdot\left(\frac{1+C_2^v}{2}\right)}\ , \qquad (3.14)$$

где $C_v=\frac{\sigma(b)}{b}$ - обозначает коэффициент вариаций времени обслуживания;

$S=\frac{1}{\mu}$- среднее время обслуживания, μ - интенсивность обслуживания;

$\rho = \frac{\lambda}{n \cdot \mu}$ - средняя загруженность системы, λ - средняя интенсивность входящих вызовов.

Для модели СМО вида *M/Ш/n* выражения для вычисления приближенного значения пребывания запроса в системе будут иметь вид

$$T=W+S, \quad (3.15)$$

$$W=\frac{\mathbf{1}}{\boldsymbol{n}} \cdot \mathbf{S} \cdot \left(\frac{\mathbf{1}}{\mathbf{1}-\boldsymbol{\rho}}\right) \cdot \left(\frac{\mathbf{1}+\boldsymbol{C}_2^v}{\mathbf{2}}\right), \quad (3.16)$$

$$S=\exp\left(\frac{1}{2} \cdot \sigma^2 + m\right),$$

где - $C_2^v = e^{\sigma^2} = 1$ - параметры логнормального распределения для процесса обслуживания,

C_2^v - квадрат коэффициента вариации процесса обслуживания.

На рисунке 3.8 представлена зависимость времени ожидания запроса операторской подсистемой от загрузки системы при показательном характере распределения интервалов поступления и логнормальном для времени обслуживания запросов, по сравнению с показательным характером распределений. Параметры распределения времен обслуживания запросов σ = 1.35, m= -0.89. Среднее время обслуживания равно 1.021 мин. Максимальная средняя интенсивность поступления запросов за минуту λ = 11.748, при этом n = 1 и $\boldsymbol{n}$ = 12. Параметры модели СМО M/M/п выбраны соответствующие.

Зависимость процесса обслуживания от содержания конкретного сервера выражается в некотором несоответствии экспериментальных данных многих систем, проанализированных, например, в работах М. Кровелла и А. Доуни. Их эксперименты указывают на соответствие процессов обслуживания запросов (распределений времен обслуживания запросов) распределению Парето (P) либо логнормальному распределению (LN). Не смотря на эти результаты, можно сказать, что в нашей работе web подсистема МКЦ нас особо не интересуют, в отличие от операторской подсистемы

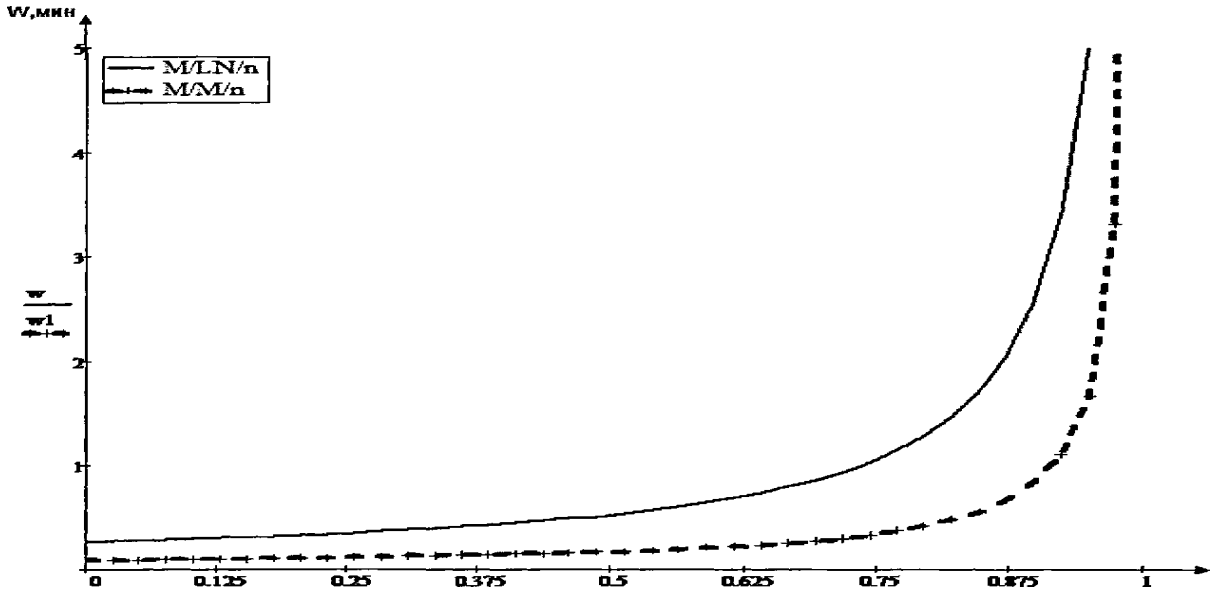


Рисунок 3.8 - Зависимость времени ожидания запроса в системе от её загрузки по СМО M/M/п по сравнению с СМО M/M/n

В процессе проведенных в третьей главе диссертационной работы исследований, проанализированы характеристики потоков, поступающих в МКЦ с отложенным обслуживанием, рассмотрен процесс обслуживания и получены следующие результаты:

1. разработан алгоритм функционирования МКЦ с отложенным обслуживанием;

2. решен вопрос моделирования ряда сложных законов распределения для процессов поступления и обслуживания. В результате определена возможность применения показательного распределения для входящих запросов и медленно затухающих распределении в процессе обслуживания МКЦ в off-line режиме;

3. разработана функциональная модель операторской подсистемы МКЦ с отложенным обслуживанием;

4. на основе функциональной модели разработана аналитическая модель для определения ВВХ операторской подсистемы МКЦ с отложенным обслуживанием, с учетом медленно затухающих свойств функций, описывающих время обслуживания. В результате получено выражение для расчета среднего времени задержки запроса в системе.

4 Имитационное моделирование и экспериментальная проверка

4.1 Методика проектирования мультисервисных контакт- центров

Проектирование контакт-центра основывается на расчете параметров согласно математическим моделям, разработанным в разделах 2 и 3. В разделе 2.4.1 работы отмечалось, что одним техническим параметрам, оказывающим влияние на качество предоставления информационных услуг пользователю и уровень затрат владельца центра и исследуемыми в работе, являются число рабочих мест операторской подсистемы контакт-центра. К ним можно еще добавить и пропускная способность канала подсистемы Web для обработки off-line запросов. Последний параметр можно пренебрег, так как если пять лет назад каналы были очень дорого, но сейчас они нечего не стоить.

Возвращаясь к рисунку. 1.5 главы 1 диссертации, с использованием результатов двух предыдущих глав работы, предлагается определенный алгоритм методики проектирования МКЦ.

Подсистема прямого обслуживания включает 13 шагов.

Шаг 1. Установить основное назначение и возможности проектируемого центра информационных услуг.

Шаг 2. Определить функциональный состав операторской подсистеме, как проектируемого объекта, указанный на рисунке 2.1.

Шаг 3. Определить набор поступающих потоков запросов и предполагаемую структуру подсистемы. Установить обозначение типов потоков запросов в соответствии по материалам раздела 2.6.

Шаг 4. Определить параметры показательных распределений интервалов времени между поступающими запросами различных типов,как $^{i}/_{\lambda}$ (i=l,2,.. .,P).

Шаг 5. Определить виды распределений времен обслуживания разнотипных запросов исходя из набора поступающих потоков. Ориентироваться на показательное.

Шаг 6. Оценить граничные значения времен ожидания обслуживания для запросов каждого из типов поступающих потоков.

Шаг 7. Задать число РМО *n* операторской подсистемы.

Шаг 8. Исходя из данных по операторам контакт-центра, определить параметры распределений времен обслуживания разнотипных запросов для

всех РМО), и найти полное среднее время обслуживания $s = \mu^{-1}$. Проверить выполнение условия (2.13), при невыполнении перейти к шагу 7.

Шаг 9. В случае реализации бесприоритетной операторской подсистемы с единым законом распределения времени обслуживания запросов воспользоваться аналитическими выражениями (1.1), (1.2), (1.3), (1.4), (1.5), (1.6) главы 1 и результатами работ из раздела диссертации 2.3. Перейти к шагу 12.

Шаг 10. В случае реализации приоритетной дисциплины обслуживания запросов, учитывающей квалификации операторов, задать приоритетную стратегию по материалам раздела 2.7., в соответствии с шагом 3.

Шаг 11. Рассчитать среднее время ожидания обслуживания и пребывания разнотипных запросов в операторской подсистеме контакт-центра. Воспользоваться аналитическими выражениями (2.4), (2.6 - 2.7), (2.10 - 2.15).

Шаг 12. Проверить полученные для всех РМО значения на соответствие принятым граничным значениям. В случае соответствия им величины завершить процесс определения параметров проектируемого контакт-центра.

Шаг 13. В случае неудовлетворительного значения величины пересмотреть предполагаемое число операторов центра *n* и/или выбранную приоритетную стратегию, затем повторить алгоритм, вернувшись к шагу 7.

Подсистема отложенного обслуживания включает 8 шагов.

Шаг 1. Установить основное назначение и возможности проектируемого центра информационных услуг.

Шаг 2. Определить функциональная структура операторской подсистеме, как проектируемого объекта, указанный на рис. 3.5.

Шаг 3. Определить параметры показательных распределений интервалов времени между поступающими потоками запросами.

Шаг 4. Определить виды распределений времен обслуживания разнотипных запросов исходя из набора поступающих потоков. Ориентироваться на показательное, детермирование, логнормальное или её ограничении соответствии по материалам раздела 3.3.

Шаг 5. Задать число п операторской подсистеме.

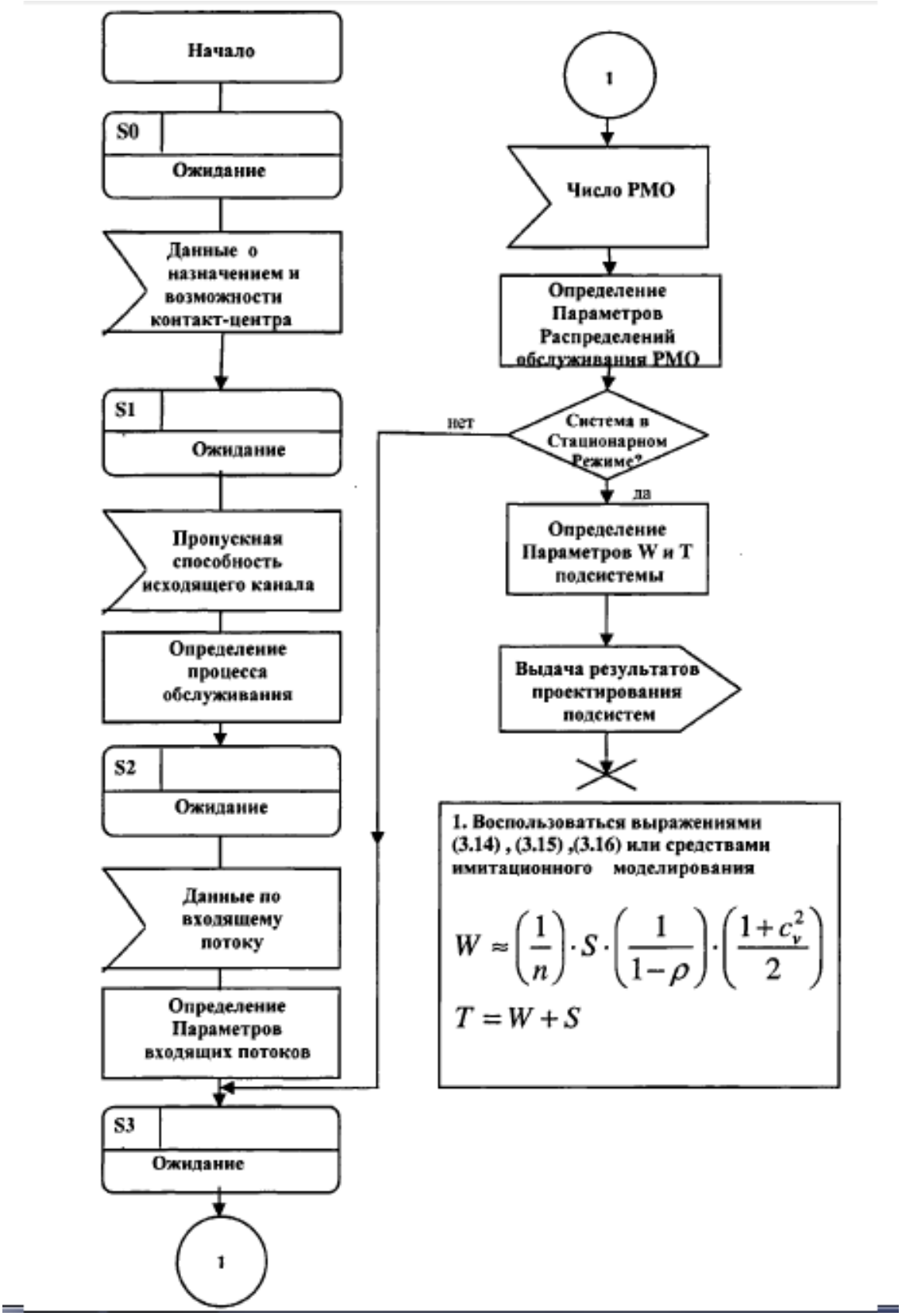
Шаг 6. Исходя из данных по операторам КЦ, определить параметры распределений времен обслуживания запросов для всех РМО, ориентироваться на логнормальное. Проверить выполнение условия стационарности системы, при невыполнении перейти к шагу 6.

Шаг 7. Рассчитать среднее время ожидания обслуживания и пребывания запросов в операторской подсистеме КЦ. Воспользоваться аналитическими выражениями (3.14), (3.15), (3.16) или разработанными средствами имитационного моделирования.

Шаг 8. В случае неудовлетворительного значения время ожидания обслуживания перейти к шагу 5.

SDL - диаграммы методики проектирования мультисервисных контакт-центров представленны на рисунках 4.1 и 4.2.

Состояния БО - БЗ и старт процесса определяют время, в течение которого процесс определения искомых параметров контакт-центра находится в режиме ожидания ввода исходных данных. Общее назначение контакт-центра, данные по входящим потокам запросов, удовлетворительное время ожидания обслуживания и необходимое число мест операторской подсистемы.



Рисуноук 4.1 - SDL-диаграмма методики проектирования on-line контакт-центров

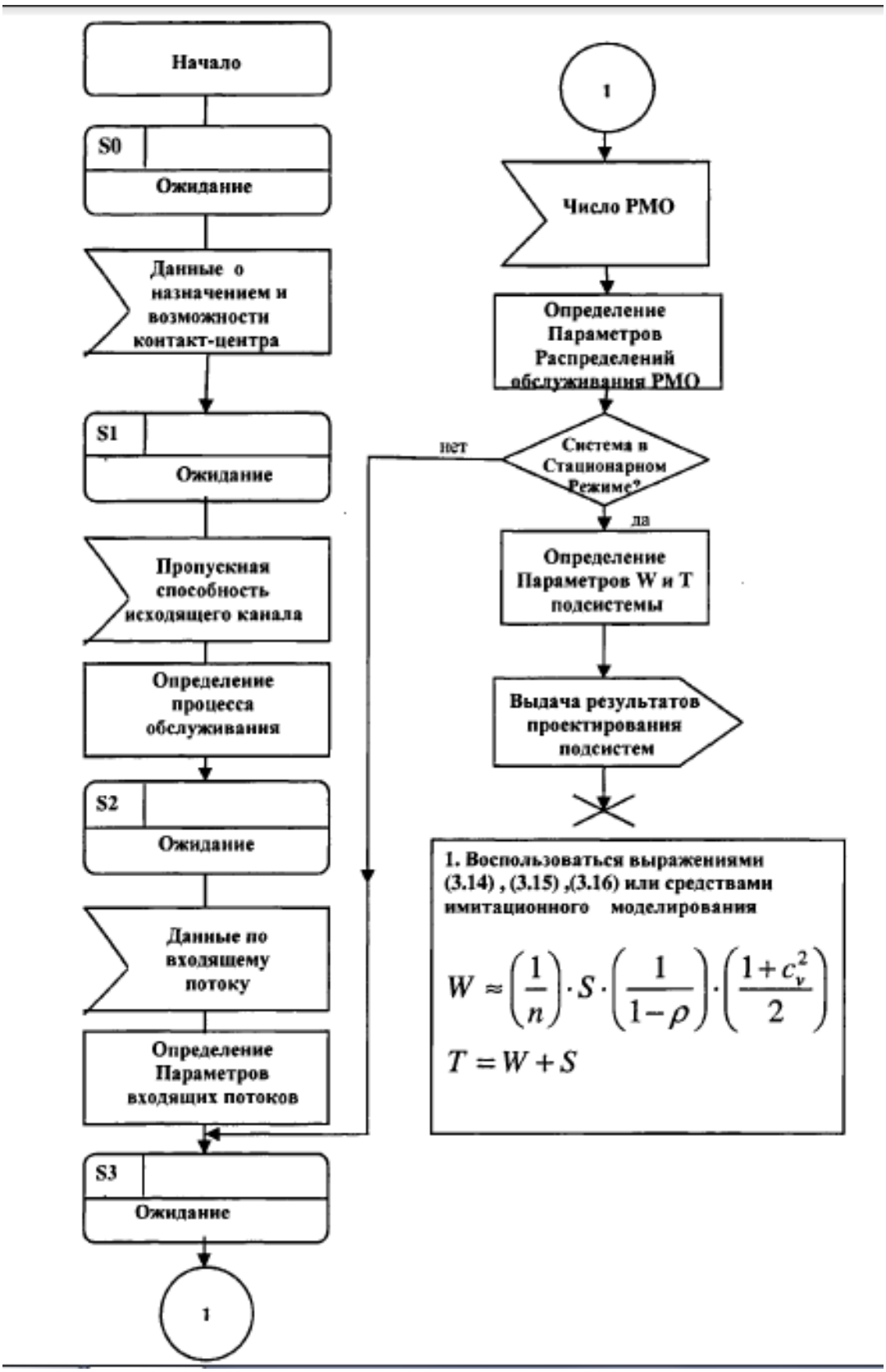


Рисунок 4.2 - SDL-диаграмма методики проектирования off-line контакт-центров

4.2 Экспериментальная проверка результатов работы на базе ситуационного контакт-центра

Экспериментальная проверка моделей, предложенных в главах 2 и 3, производилась в два шага. Это статистическое моделирование с применением языка моделирования GPSS и натурный эксперимент.

Базой для натурного эксперимента в целях проверки предложенной во второй главе работы модели является практическая реализация комплекса контакт-центра

Цель эксперимента является нахождение числа операторов, необходимых в каждой смене (7 дней работы), которая удовлетворяла бы кадровые потребности предложенных моделей массового обслуживания в главе 2. Результат эксперимента представлен на рисунке 4.3

В МКЦ поступают 3 разных типов вызовов: первый - экстренные вызовы милиции (02), второе - комплексный экстренный вызов - это одновременный вызов милиции, скорой помощи или пожарной части (02, 03, 01), третий - не экстренные вызовы, т.е. обычные вызовы. Далее эти 3 вида вызовов разделены на 2 приоритетных класса (P=2): P_1 - экстренные вызовы (02) с возможностью одновременного вызова (03) и (01), P_2 - обычные вызовы.

Предложенные в работе математические модели позволяют получить необходимое количество операторов для часовых временных интервалов типовой рабочей недели (168 часов, с полуночи в понедельник к полуночи в воскресенье), которое обеспечит работу полицейского контакт-центра и обслуживание экстренных и не экстренных запросов с заданным качеством. Это изменение может быть связано с эффект временем суток.

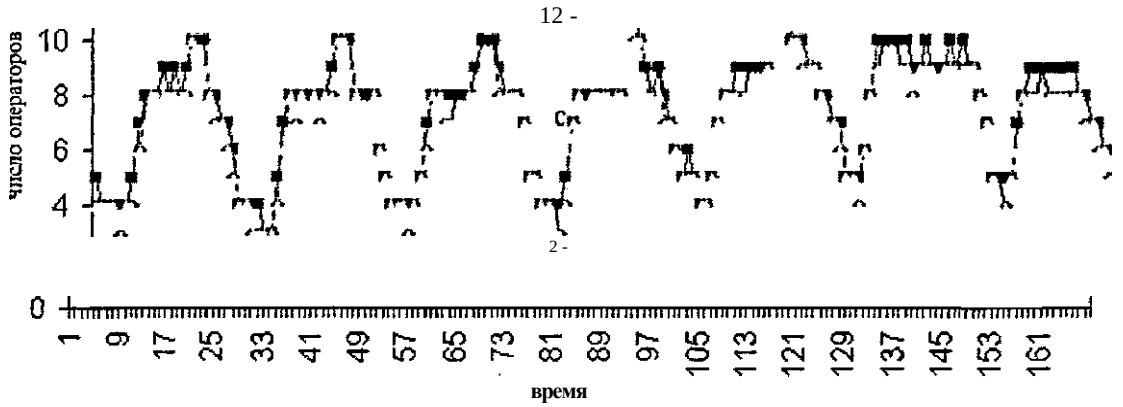


Рисунок 4.3 - Зависимость распределение число операторов в системе от 7 дней

В частности, наблюдается, что число персонала достигла своего пика примерно около период с 4 до 7 вечера в будние дни, тогда как в выходные штатное расписание предсказал менее переменной. Подозревается, что

эффект время суток играет важную роль в будние дни. Стандартная модель M/M/n.

На самом деле в МКЦ больше внимание выделяется запросы приоритет 1, так как запросы P_2 менее важные, их можно игнорировать. Следовательно, действуя по методике раздела 4.1, воспользуемся выражением (1.3) главы 1 и (2.1) глава 2 , где определяются число n операторы.

$$n \approx A + \beta\sqrt{A}, \tag{3.17}$$

где A -нагрузка на МКЦ;

β- параметр качества обслуживания.

Значение β получается из предела выражения для n, и интенсивности поступающих вызовов $A = n - \rho - 4\sqrt{\boldsymbol{n}}$.

Для стабильности системы $\rho < 1$, тогда минимальное число операторы равно.

$\mathbf{n_{min}} = \boldsymbol{\lambda}/\boldsymbol{\mu} \approx \mathbf{4}$ и $n \approx \mathbf{A} + \boldsymbol{\beta}\sqrt{\mathbf{A}} \approx \mathbf{12}$, при β=2, λ=0,042, $\boldsymbol{\mu} = \mathbf{1}/\mathbf{90}$ ρ<1

Проведенный расчет число оператора совпадает с экспериментальными данными. Таким образом, для рассмотренных подсистем предполагаемого контакт-центра могут быть сделаны следующие выводы:

1. число операторов полученные в результате расчета оказывается несколько завышенным по отношению к данным эксперимента, если растет суммарную загрузку A, но все, же достаточны для приближенной оценки на практике;

2. теоретически, штатное расписание предсказано, в приоритетны модели, должна быть меньше, чем у стандартной модели. Это связано с тем, что все звонки обрабатываются как приоритет 1 в стандартной модели, в то время как в приоритетных модели, звонки делятся на приоритет 1 и приоритет 2 , у которого более низкую производительность;

3. оценка используемого комплекса технических средств проводилась во времени эксперимента достаточно для выяснения всех аспектов функционирования ситуационного центра. Можно констатировать, что переход на 1Р-технологии позволил обеспечить требования, предъявляемые к качеству передачи информации (речи, данных и видео) при связи с операторами экстренных служб. При этом следует учесть, что расчет числа операторов должен быть сделан корректно при наличии их квалификации, должен быть правильно осуществлен выбор математической модели. Тогда не будет наблюдаться перегрузка системы.

В случае произвольных потоков заявок, отличающих от простейших, значительно усложняется аналитический расчет характеристик обслуживания заявок, что предопределяет применение имитационного моделирования.

Проверка результатов главы 3 выполнялась средствам имитационное моделирования в разделе 4.3.

4.3 Имитационная модель МКЦ с отложенным облуживанием вызовами на ОРББ

В математических моделях (ММ) сложных объектов, представленных в виде систем массового обслуживания (СМО), фигурируют средства обслуживания, называемые обслуживающими приборами (ОП), и обслуживаемые заявки.

Состояние СМО характеризуется состояниями ОП, заявки и очередей к ОП. Состояние ОП описывается двоичной переменной, которая может принимать значения "занят" или "свободен". Переменная, характеризующая состояние заявка, может иметь значения "обслуживания" или "ожидания". Состояние очереди характеризуется количеством находящихся в ней заявок.

Особым классом математических моделей являются имитационные модели. Они представляют собой компьютерную программу, которая шаг за шагом воспроизводит события, происходящие в реальной системе. Имитационная модель СМО представляет собой как программная модель сложной системы, в которой отражены структура, алгоритмы развития и протекания процессов во времени, временные характеристики отдельных элементов. Имитация имеет своей основной целью моделирование динамики, т. е. изменение состояния системы во времени. Как и всякому формализованному подходу, имитационному моделированию присущи свои понятия и атрибуты. *Время моделирования* - это временной интервал, на котором имитируется поведение системы, т.е. в сущности, имитационное представление реального времени.

Активность - наименьшая единица работы при выбранном уровне представления моделируемой системы, которая рассматривается как единый дискретный шаг.

Процесс - совокупность логически связанных активностей.

Событие - мгновенное изменение состояния некоторого объекта системы,

который может быть активным либо пассивным. События можно разделить на две категории: события следования (управляют инициализацией активностей внутри данного процесса) и события изменения состояния (управляют выполнением активностей, относящихся в общем случае, к независимым процессам).

Транзакты - динамические объекты, представляющие собой поток элементов обслуживания и являющиеся конкретной реализацией процессов. Функционально-ориентированные объекты, которые соответствуют элементам оборудования или рабочим, обслуживающим транзакты.

Очереди можно рассматривать как статические объекты, позволяющие оценить поведение системы. В очередь попадают транзакты, которые

задержаны в какой-то момент времени до тех пор, пока не выполнится условие, необходимое для их продвижения. Необходимо отметить, что при моделировании процессы, события и активности целиком зависят от потоков и траекторий движения транзактов: транзакт, попадая в моделируемую систему, занимает определенные блоки, вызывая при этом события. Наступление событий должно планироваться соответствующими средствами моделирования. При выполнении определенных условий событие вычеркивается из системы моделирования, а на смену ему должны приходить следующие события. При событийном моделировании производственной системы выделяют узловые моменты динамики в виде событий. В процессе моделирования осуществляется переход (скачок во времени) от предыдущего события к последующему. Каждое событие выполняется мгновенно во времени, модельное время затрачивается только на переход от события к событию. Реализация событий во времени напоминает цепную реакцию при отработке любого события планируется одно или несколько последующих (будущих) событий.

При имитационном моделировании определяются входные и выходные потоки. Параметры входных потоков заявок – это внешние параметры СМО.

Выходными параметрами являются величины, характеризующие свойства системы - качество ее функционирования. Примеры выходных параметров, производительность СМО - среднее число заявок, обслуживаемых в единиц времени; коэффициенты загрузки оборудования - отношение времен обслуживания к общему времени в каждом ОП; среднее время обслуживания одной заявки. Основное свойство ОП, учитываемое в модели СМО, - это затраты времени на обслуживание, поэтому внутренними параметрами в модели СМО являются величины, характеризующие это свойство ОП. Обычно время обслуживания рассматривается как случайная величина и в качестве внутренних параметров фигурируют параметры законов распределения этой величины.

Имитационное моделирование позволяет исследовать СМО при различных типах входных потоков и интенсивностях поступления заявок на входы, при вариациях параметров ОП, при различных дисциплинах обслуживания заявок. Основной тип ОП - устройства, именно в них происходит обработка транзактов с затратами времени. К ОП относятся также накопители (памяти), отображающие средства хранения обрабатываемых деталей в производственных линиях или обрабатываемых данных в вычислительных системах. Накопители характеризуются не временами обслуживания заявок, а емкостью - максимально возможным количеством одновременно находящихся в накопителе заявок. К элементам имитационных моделей СМО кроме ОП, относят также узлы и источники заявок. Связи ОП между собой реализуют узлы, т.е. характеризуют правила, по которым заявки направляются к тому или иному ОП.

Преимуществом имитационных моделей является возможность подмены процесса смены событий в исследуемой системе в реальном масштабе

времени на ускоренный процесс смены событий. В результате можно воспроизвести работу системы в течение продолжительного времени, что дает возможность оценить её работу в широком диапазоне варьируемых параметров.

При реализации имитационного моделирования на ЭВМ производится накопление статистических данных по тем атрибутам модели, характеристики которых являются предметом исследований. По окончании моделирования накопленная статистика обрабатывается, и результаты моделирования получаются в виде выборочных распределений исследуемых величин или их выборочных моментов. Таким образом, при имитационном моделировании СМО речь всегда идет о статистическом имитационном моделировании.

Для описания моделей СМО при их исследовании на ЭВМ разработаны специальные языки имитационного моделирования. Существуют общецелевые языки, ориентированные на описание широкого класса СМО в различных предметных областях, и специализированные языки, предназначенные для анализа систем определенного типа. Примером общецелевых языков служит широко распространенный язык GPSS.

На персональных компьютерах типа IBM/PC язык GPSS реализован в рамках различных пакетов прикладных программ. В целях диссертационной работы применялся пакет GPSS World Student Version. В диссертационной работе методы имитационного моделирования применялись в целях проверки аналитических выражений, полученных в главы 3, а также как основное средство решения задач работы в тех случаях, когда получение аналитических выражений было затруднено.

Имитационное моделирование операторской подсистемы с отложенным обслуживанием, включающее показательные законы поступления, распределения с тяжелым хвостом (логнормальные) законы обслуживания, проведено в четвертой главе. Особенность программы заключается в том, что она позволяет легко добавлять и удалять нужные поступающие потоки нагрузки, в зависимости от исследуемой системы. Реализация распределений процессов обслуживания запросов в данной системе основана на встроенных в программный пакет GPSS функциях. Набор текста программ, разработанных в рамках диссертации, представлен в приложении к работе. Рис. 4.4 представляет зависимость времени ожидания (Имитационное моделирование) запросы операторской подсистемы от загрузки системы для СМО M/LN/n по сравнению с результатом приближения аналитическим выражений для СМО M/LN/n. Параметры указанных моделей для процесса обслуживания выбраны таким же, как для расчета зависимостей рис. 3.8. главы

В данном случае, на уровне загрузке 0,875 расхождение результатов аналитического и имитационного моделирования, составляет около 10%.

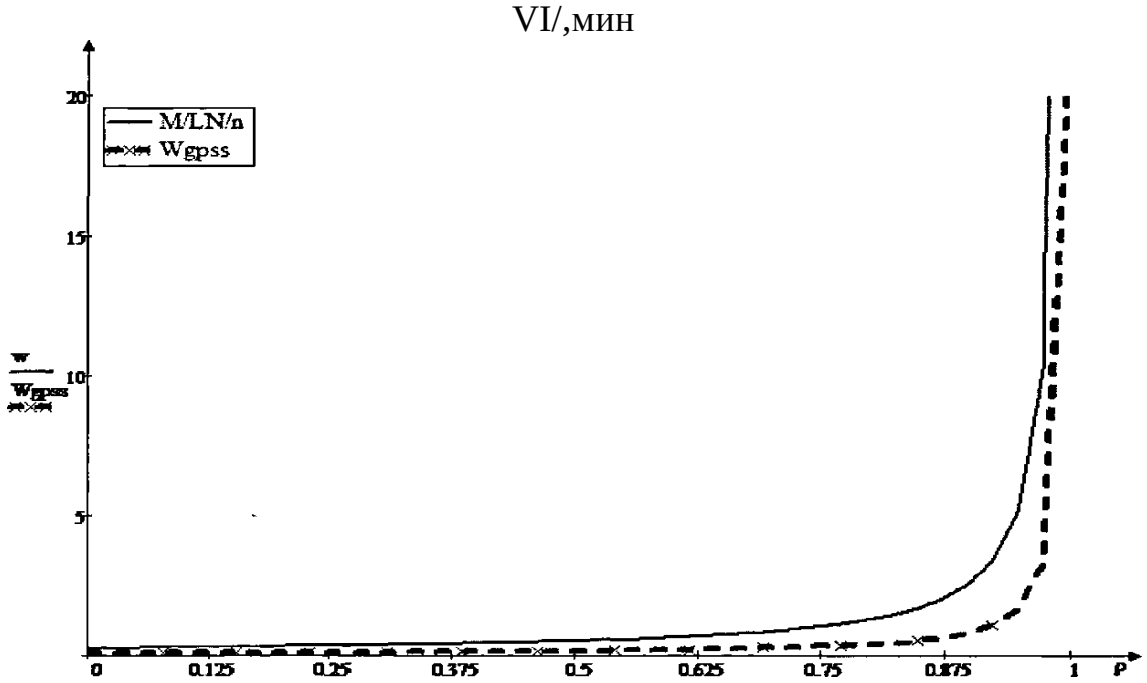


Рисунок 4.4 - Зависимость времени ожидания запроса в системе СМО M/LN/n от ёе загрузки и результаты имитационного моделирования

По материалам четвертой главы работы, посвященной экспериментальной проверке моделей предложенных в 2 и 3 главах и практические реализации современного МКЦ, сформулированы следующие выводы:

1. с использованием математических моделей и методов расчёта ВВХ, предложенных в 2 и 3 главах диссертационной работы, разработан алгоритм проектирования мультисервисных контакт-центров;

2. предложенная методика проектирования опробована на существующей реализации ситуационного контакт-центра в качестве натурного эксперимента;

– разработана имитационная модель для определения ВВХ операторской подсистемы контакт-центра с отложенным обслуживанием.

Заключение

Данная магистерская диссертация посвящена исследованию моделей и методов расчета мультисервисных контакт-центров.

В работе освещены исследование специфики процессов обслуживания запросов в МЦК. Выбран метод расчетов контакт-центра при обслуживании разнотипных потоков вызовов по приоритетной дисциплине.

Определено основные параметры влияющие на качество предоставления информационных услуг.

В первой главе проведен анализ современного состояния исследований и анализа трафика мультисервисного контакт - центра. Рассмотрены вопросы эволюции Центров обслуживания вызовов, назначение и облости применения современных контакт центров. Проведен статистический анализ неравномерности по часам сутки. В работе рассматривается возможность применения методов экстраполяции и скользящей средней для определения прогнозных оценок входящего трафика контакт - центра горизонтам трафика 1-3 дня. Предложенная методика прогнозирования может быть использована в аналогичных контакт- центрах операторов связи для обеспечения качественного обслуживания клиентов.

Список литературы

1. Соколов Н.А. Телекоммуникационные сети. - М.: Альварес Паблишинг, 2003.-196 с.

2. Кучерявый А.Е., Парамонов А.И, Кучерявый Е.А. Сети связи общего пользования. Тенденции развития и методы расчета. М.: ФГУП ЦНИИС, 2008, - 232с

3. Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. Монография /Под ред. О.И.Шелухина - М.: Радиотехника, 2003., 198 с.

4. http://ru.wikipedia.org/wiki/PON

5. http://www.skomplekt.com/tovar/1/0/pon/

6. http://www.eltex.kz/model-postroeniya-seti-pon-predostavlenie-uslugi-na-port

7. Кучерявый А.Е., Пяттаев В.О. Новое терминальное оборудование для пакетных сетей. /Тез. докладов международной конференции «Развитие услуг связи на основе телекоммуникационных технологий нового поколения (NGN-2003)» С. Петербург 2003, с. 55-56.

8. Никитин А.В., Пяттаев В.О., Никульский И.Е., Филиппов А. А. Концепция построения мультисервисной сети оператора связи. //Вестник связи. 2010 №5. - с. 47-49, №7. - с. 41-45.

9. Соколов Н.А. Сети абонентского доступа. Принципы построения. //ЗАО «ИГ» «Энтср-профи», 1999.

10. Никитин А.В., Микульский П.П., Филиппов А.А. Особенности внедрения технологий PON на сети оператора, занимающего существенные рыночные позиции. // Вестник связи. 2009, №8.- с.7-9.

11 ITU-T. Recommendation G.107. The E-model, a compytational model for Use intransmission planning.

13. ETSI. Telecommunications and Internet Protocol Harmonization over Networks (TIPHON) Relase 3; End-to-End Qualiti of Service in TIPHON Systems; Part 7: Design guide for elements of a THIPHON connection from end to end speech transmission performance point of view. -TR 101 329-7,2002.

14. Гроднев, И.И. Мурадян А.Г., Шэрафутдннов Р.М. и др. Волоконно-оптические системы передачи и кабели. Справочник - М.: Радио и Связь, 1993.- 264 с.

15. ITU-T Recommendation Y.1541 (02/2006) - Network performance objectives for IP-based services.

16. Рекомендация МСЭ-Т Y.1541 (2006 г.) изменение 1: Новое дополнение X-Пример, показывающий метод расчета IPDV на основе множества сегментов.

17. Вадзинский Р.Н. Справочник по вероятностным распределениям. СПб.: Наука, 2001.

18. Кучерявый Е.А. Управление трафиком и качество обслуживания в сети Интернет. - СПб.: Наука и техника, 2004.

19. Крылов В.В., Самохвалова С.С. Теория телеграфика и ее приложения. - СПб.: БХВ - Петербург, 2005.

20. Захаров Г.П. Методы исследования сетей передачи данных. М.: Радио и связь. 1982.

21. Нейман В.II. Самоподобные процессы п их применение в теории телетрафика. //Труды МАС. 1999. - № 1(9).- с. 11-15.

22. Галкин А.М., Симонина О.А. Метод расчета характеристик IP-ориентированных мультисервисных сетей с учетом свойств самоподобия трафика. //Труды учебных заведений связи. СПб.: 2005.

23. Горелов Г.В. О применимости методов классической теории телетрафика в исследованиях систем с пакетной коммутацией. Ведомственные и корпоративные сети и системы (ВКСС), 2006, №1 с. 103-107.

24. Вишневский В.М. Теоретические основы проектирования компьютерных сетей. М.:Техносфера, 2003

25. Алиев Т.Н. Основы моделирования дискретных систем. - СПб, СПбГУ ИТМО, 2009.

26. Kramer W.. Langenbach - Beiz M. Approximate formulae for the delay in the queuing system G/G/l /V'Congressbook, S'" Intranet. Telemiffic Congress, Melbome (1976).

27. Суздалев А.В., Чугреев О.С. Передача данных в локальных сетях связи.- М.: Радио и связь, 1987.

28. Симонина О.А., Яновский Г.Г. Характеристики трафика в сетях IP. //Труды учебных заведений связи. СПб., 2004, с.8-14.

29. Боев В.Д. Моделирование систем. Инструментальные средства GPSS World Учебное пособие, БХВ-Петербург, 2004. - 368 с.

30. Учебное пособие по GPSS World. – Казань: Изд-во «Мастер Лайн», 2002.

31. Руководство Пользователя по GPSS World. – Казань: Изд-во «Мастер Лайн», 2002.