

СЕВЕРО-КАВКАЗСКИЙ ФИЛИАЛ
ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«Московский технический университет связи и информатики»

Кафедра «Инфокоммуникационные технологии и системы связи»

НЕРСЕСЯНЦ А.А.

Теория телетрафика

Учебное пособие по дисциплине

ТЕОРИЯ ТЕЛЕТРАФИКА

Направление подготовки 11.03.02
«Инфокоммуникационные технологии и системы связи»,
профиль «Сети связи и системы коммутации»

Ростов-на-Дону
2013 год

Учебное пособие по дисциплине ТЕОРИЯ ТЕЛЕТРАФИКА.

Направление подготовки 11.03.02
«Инфокоммуникационные технологии и системы связи»,
профиль «Сети связи и системы коммутации»

Методическое пособие по Теории телетрафика предназначено для изучения студентами очной и заочной форм обучения основных положений этой теории применительно к современным мультисервисным сетям и, в первую очередь, к IP-сетям с коммутацией пакетов.

Автор:
профессор кафедры ИТСС,
д.т.н., с.н.с. Нерсисянц А.А.,

Рассмотрено и одобрено на заседании кафедры СССК
Протокол №11 от 27.06.2016

ISBN 978-5-904033-05-7



9 785904 033057

Технический редактор: Григораш О.И.

И з д а т е л ь с т в о С К Ф М Т У С И

Сдано в набор 19.02.13. Изд. № 32. Подписано в печать 21.02.13. Зак. № 45.

Печ. листов Учетно-изд. л..... Печать оперативная. Тир. 160 экз.

Отпечатано в Полиграфическом центре СКФ МТУСИ, Серафимовича, 62.

Содержание

Введение.....	4
1. Предмет теории телетрафика.....	4
2. Место теории телетрафика в методах анализа ТК-систем.....	6
3. Базовые понятия теории телетрафика.....	8
4. Потоки вызовов.....	9
5. Характеристики качества обслуживания.....	19
6. Дисциплины обслуживания.....	21
7. Нагрузка, работа и их виды.....	24
8. Системы с явными потерями.....	26
9. Системы с ожиданием.....	31
10. Системы с повторными вызовами.....	35
11. Управляемые элементы сети связи и методы определения ее характеристик.....	36
12. Анализ буферной памяти узла коммутации.....	42
13. Расчёт характеристик узлов коммутации в сетях.....	48
14. Анализ мультисервисных сетей.....	53
15. Управление потоками в сети с пакетной коммутацией.....	59
16. Расчёт числа каналов в сотовых сетях.....	64
17. Исследование телекоммуникационных систем и сетей методом имитационного моделирования.....	66
Приложение.....	68
Список использованной литературы.....	86

Введение

Теория телетрафика – как самостоятельная дисциплина выделилась из теории массового обслуживания и развивалась в течении нескольких десятилетий почти исключительно для нужд исследования и расчёта телефонного трафика. Появление трафика данных во второй половине прошлого века не могло существенно изменить это положение, поскольку объёмы этого трафика были пренебрежимо малы.

Показательно в этом плане, что возникшая в 80-х годах технология **ISDN** (*Integrated Services Digital Network*) — цифровая сеть с интеграцией служб, предназначенная для совмещения услуг телефонной связи и обмена данными, рассматривала трафик данных как нечто несущественное. Именно по этой причине интегрированный трафик базировался, в основном, на “суперскоростных” цифровых каналах в 64 кбит/с и предлагал компьютерам использовать для обмена бесперспективный способ коммутации каналов. В конечном счёте, именно эти два фактора и определили печальную судьбу этой технологии на границе веков (максимальная доля трафика ISDN нигде не превышала 5%).

Оставив в стороне ISDN, обратим внимание на стремительное ускорение пакетных технологий в фиксированных (X.25, Frame Relay, ATM, MPLS) и мобильных (GPRS, EDGE, WCDMA, HSPA, WiMAX, LTE) сетях. Закон Мура торжествует как везде.

Какие аспекты в этом развитии оказали наибольшее влияние на формирование новой парадигмы теории телетрафика?

В первую очередь это входной поток. Благоденствие без последствий вдруг закончилось и на смену идеальных во всех отношениях пуассоновских потоков пришли понятия самоподобного трафика, пачечной нагрузки и др.

Следующий не менее важный аспект связан с дисциплиной обслуживания. Услуги телефонной связи (особенно автоматической) при способе коммутации каналов предоставлялись почти исключительно по *системе с явными потерями*. Сама система АТС и не предполагала каких-либо мест для хранения фрагментов речевой информации. В то же время, сам принцип коммутации пакетов (прими – сохрани – передай дальше) предполагает организацию очередей в каждом узле коммутации и, следовательно, обслуживание по *системе с неявными потерями*.

Далее, рост пропускной способности каналов. В последние годы сертифицированы технологии Ethernet в 40 и 100 Гбит/с на одну спектральную составляющую. И дело здесь не только в количественном росте. В эпоху “чистой” телефонии какой бы большой ни была скорость физического канала, каждому соединению предоставлялся в *монопольное владение* небольшой ресурс этого канала. При коммутации пакетов даже в режиме виртуального канала в распоряжение передаваемого пакета на небольшой промежуток времени предоставляется весь канальный ресурс. Отсюда, меняется один из главных параметров исследуемых в задачах теории телетрафика. Вместо ана-

лиза многоканальной системы и определения числа каналов, необходимых для передачи заданного трафика при требуемой вероятности явных потерь, в современных сетях основной характеристикой становится вероятность передачи пакета по одноканальной системе с задержкой не хуже допустимой (разумеется с учётом ожидания в очереди). Т.е. налицо кардинальное изменение целевой функции расчётной схемы.

И наконец, измерение объёма трафика (потoki, тяготения). Конечно Эрланги, или ТЧ-каналы, или число 30-канальных групп, т.е. потоков Е1 (в простонародье «двушек»). Но всё это не для Internet и в системы All-IP с такими единицами измерений не войдёшь. Для мультисервисных сетей с совокупным контентом Triple Play (A, V, D) нужна единая, удобная для всех видов потоков единица, например, байт/с (оставим бит/с как всегда для физических каналов). Можно оперировать и потоками пакетов/с, но только, если в рассматриваемой системе длины пакетов подчинены одному и тому же распределению. Теперь любые потоки легко складываются и разделяются, но речь — только как VoIP.

Анализу современных проблем в теории телетрафика посвящён целый ряд монографий [2, 4, 5, 6, 7, 8, 11]. Сложность поставленной задачи (речь, как правило, идёт об анализе мультисервисных систем) предопределила и математическую сложность полученных решений, большая часть которых заинтересует научных работников, в том числе и соискателей учёных степеней. Однако, для студентов инженерных факультетов и особенно для студентов заочного обучения, во избежание формульной перенасыщенности, приходится тщательно фильтровать включаемый в лекционный курс материал. Естественно, что фильтрация упрощает задачу, а упрощение достигается введением многих ограничений, которые отдаляют построенную аналитическую модель системы от реальной системы. С этим отдалением приходится мириться ради доступности преподаваемого материала.

В настоящее пособие включены материалы, использованные автором в учебном процессе в течении нескольких лет при проведении занятий по дисциплине Теория телетрафика на очных и заочных факультетах Северо-Кавказского Филиала Московского Технического Университета Связи и Информатики.

1. Предмет теории телетрафика

Теория телетрафика (ТТ) научная дисциплина, предназначенная для изучения процессов обслуживания системами распределения информации поступающих потоков сообщений и количественная оценка качества обслуживания.

Математическая модель процесса обслуживания включает четыре основных элемента:

- система распределения информации (система обслуживания – узлы, линии связи, сети);
- входной поток сообщений;
- дисциплина обслуживания;
- характеристики качества обслуживания.

ТТ решает следующие задачи:

- исследование входного потока вызовов и описание его характеристик;
- исследование и нахождение оптимальной структуры системы обслуживания;
- исследование и определение оптимальных алгоритмов обслуживания вызовов;
- определение пропускных способностей структурных элементов системы (узлов и линий связи);
- определение качества обслуживания вызовов.

Доминирующее положение в ТТ занимают задачи анализа – определение характеристик качества обслуживания в зависимости от параметров и свойств потока вызовов, параметров и дисциплины системы обслуживания. Более сложной является обратная задача – задача синтеза, т.е. нахождение параметров системы в зависимости от свойств входного потока и требуемого качества обслуживания.

2. Место теории телетрафика в методах анализа ТК-систем

Для исследования процессов в различных реальных системах, как правило, создаются модели, отражающие свойства системы, существенные для целей данного анализа. Например, для изучения зависимостей между напряжением, током и сопротивлением используется аналитическая модель – закон Ома ($U = IR$).

Наиболее популярные модели, используемые для изучения телекоммуникационных систем и различающиеся формой отображения объекта, представлены на рис. 2.1.

При натурном моделировании характеристики исследуются на реальном объекте. Модель адекватна, но возможности ограничены. Нельзя, например, для целей эксперимента попросить абонентов звонить чаще или говорить меньше.

При математическом моделировании реальному объекту устанавливается соответствие некоторого математического объекта – модели. В зависимости от сложности модели создаётся та или иная степень приближения модели к реальному объекту. Различают аналитическое и имитационное математическое моделирование.

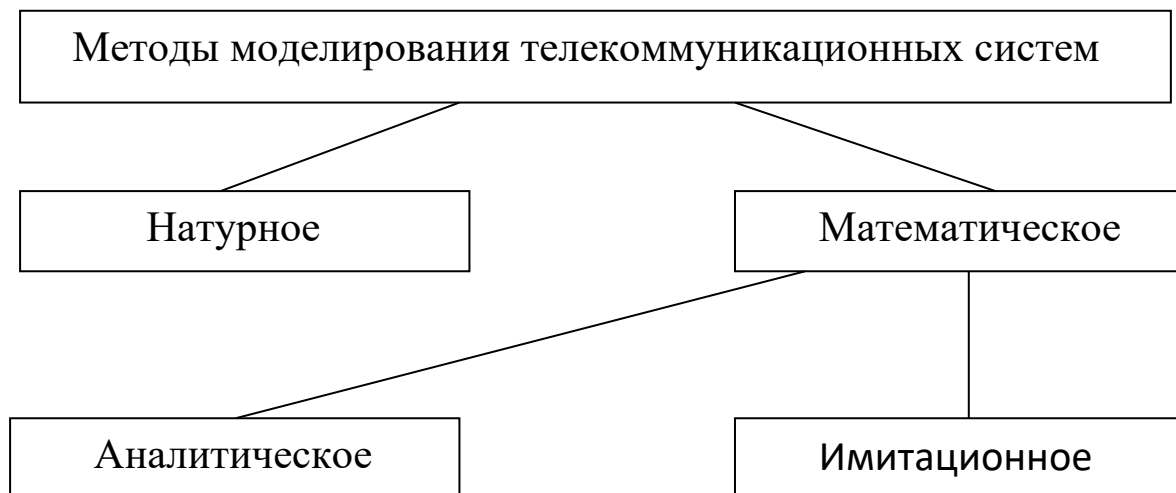


Рис. 2.1. Основные виды моделирования

Аналитическое – процесс функционирования системы отображается функциональными соотношениями (уравнениями) или логическими условиями. Уравнения решаются в общем виде или численными методами.

Имитационное – процесс функционирования системы воспроизводится с помощью моделирующего алгоритма – программы для ЭВМ. При многократном воспроизведении случайных событий, величин и функций и последующей статистической обработке результатов определяются статистические характеристики объекта исследования. Практически неограниченный класс исследуемых систем.

По сравнению с аналитическим моделированием имитационное позволяет получить сколь угодно точное отображение реального объекта, правда за счёт времени и денег, затраченных на разработку моделирующей программы, и машинного времени расходуемого на прогоны. Иногда на это уходят дни и месяцы.

Аналитическое моделирование при наличии предварительно разработанного математического аппарата позволяют легко получать нужные характеристики. Однако аналитическому описанию поддаётся только очень ограниченный круг систем. В основном, это системы без последствий. Поэтому при выборе подходящей аналитической модели из имеющегося ряда приходится делать допущения о характере реальных процессов, что, в конечном счёте, делает модель менее адекватной.

Теория телетрафика – исследование телекоммуникационных систем методом аналитического моделирования.

3. Базовые понятия теории телетрафика

Сообщение – это форма представления информации, имеющая признаки начала и конца, и предназначенная для передачи через сеть или коммутационную систему. Сообщение характеризуется типом, объёмом, формой представления информации, адресами отправителя и получателя и другими атрибутами, необходимыми для его продвижения по сети.

Примеры сообщений – IP-пакет, телефонный разговор, телеграмма, факсимильное сообщение, ТВ-сигнал и т.д.

В теории телетрафика различают следующие виды сообщений:

- обслуженное – переданное через сеть связи или коммутационную систему (КС);
- потерянное – поступившее в сеть связи, но не переданное полностью получателю по разным причинам (занятость или недоступность соединительных линий, повреждения, и т.д.);
- задержанное – поступившее в сеть и ожидающее начала передачи;
- условно потерянное – поступившее в сеть и задержанное сверх допустимого времени.

Вызов – требование источника на установление соединения, поступившее в сеть связи, КС, коммутационное оборудование, с целью передачи или обслуживания сообщения.

Необходимо отметить, что вызов, в отличие от сообщения, характеризуется только моментом поступления и не имеет длительности.

В качестве источников вызовов могут выступать персональный компьютер, телефонный, телеграфный или факсимильный аппарат, абонентский терминал, управляющее устройство. Часто одни и те же устройства могут служить и источниками, и приемниками вызовов.

Виды вызовов:

- обслуженный – получивший соединение. Различают полностью обслуженный вызов (получивший соединение с требуемым приемником) и частично обслуженный вызов (получивший соединение на участке сети);
- успешный – окончившийся передачей сообщения приемнику;
- потерянный – получивший отказ в установлении соединения. Вызовы, получившие ошибочные соединения, также относятся к потерянным.
- задержанный – ожидающий начала установления соединения.
- первичный – первый для данного сообщения вызов;
- повторный (вторичный) – поступивший в сеть связи через некоторый промежуток времени после того, как был потерян предыдущий вызов, соответствующий тому же сообщению.

Занятие – любое использование прибора, линии, устройства с целью установления соединения независимо от того, закончилось оно передачей сообщения или нет.

Занятие характеризуется моментом занятия и длительностью занятия.

Освобождение – возвращение прибора, линии, устройства в исходное состояние. Освобождение характеризуется только моментом наступления.

С учетом данных определений можно определить первое базовое понятие теории телетрафика – поток вызовов.

Поток вызовов – это множество последовательных моментов поступления вызовов.

Аналогично можно определить потоки сообщений, занятий, освобождений или других однородных событий.

4. Потоки вызовов

Под потоком вызовов понимают последовательность моментов поступления вызовов на обслуживание сообщений: $t_1, t_2, \dots, t_k, \dots, t_n$

Поток вызовов называется детерминированным, если последовательность моментов поступления вызовов заранее определена и известна, и случайным, если эта последовательность случайна.

Случайные потоки вызовов характеризуются случайными интервалами между моментами их возникновения. Случайные потоки задаются вероятностными законами распределения.

4.1. Свойства и характеристики потоков вызовов

Случайные потоки вызовов классифицируются в зависимости от наличия или отсутствия трёх следующих свойств: стационарность, ординарность, последствие.

Поток вызовов называется **стационарным**, если с течением времени его вероятностные характеристики не изменяются.

Наиболее часто в качестве такой вероятностной характеристики используется величина $p_k(t)$ - вероятность поступления k вызовов за время t .

Для стационарного потока $p_k(t)$ зависит только от величины интервала t и не зависит от того, где находится этот интервал.

Реальный поток вызовов имеет, как правило, явно выраженный нестационарный характер. Наблюдаются суточные, недельные и годовые колебания. Однако внутри суток обычно можно выделить относительно небольшие промежутки времени, в течение которых поток можно считать стационарным.

Последствие означает зависимость вероятностных характеристик потока от предыдущих событий. Отсутствие последствия означает, что прошлые события (моменты поступления вызовов до момента t и процесс их обслуживания) никак не влияют на поступление вызовов после момента t .

Другими словами, поток считается потоком с последствием, если вероятность поступления вызовов в интервале $[t_1, t_2)$ зависит от числа, времени поступления и длительности обслуживания вызовов до момента t_1 .

Таким образом, для потока без последствия условная вероятность поступления k вызовов в интервале $[t_1, t_2)$, равна безусловной. Поэтому такой

поток можно задавать семейством безусловных вероятностей $p_i(t_1, t_2)$ поступления i вызовов в интервале $[t_1, t_2)$.

Стационарный поток без последействия можно задавать семейством вероятностей $p_i(t)$ поступления i вызовов в любом интервале длиной t .

Если поток поступает от большой группы источников, то такой поток близок к потоку без последействия. Если число источников мало, то поток имеет явно выраженное последействие.

Например, предположим, что число источников $N = 10$. Если все они свободны, имеем некоторое значение вероятности $p_k^{10}(t)$. Если же некоторое число источников занято, например, $N_{св} = 6$, то $p_k^6(t) < p_k^{10}(t)$. Число свободных источников зависит от предыдущих событий (от количества вызовов в предыдущие моменты времени, а, следовательно, от числа источников, способных генерировать новые вызовы). Таким образом, в этом случае имеет место явно выраженное последействие.

С ростом общего числа источников постепенно уменьшается доля занятых источников по отношению к общему их числу, соответственно, ослабевает последействие потока, и при каком-то большом значении N его можно не учитывать.

Понятие большое или малое количество источников нуждается в уточнении. Например, 300 источников по отношению к 10 каналам можно условно считать бесконечно большим, а по отношению к 200-м каналам – соизмеримым. Тогда в общем виде можно говорить об отсутствии последействия, если число источников N много больше числа каналов v , т.е. если $N \gg v$.

Потоки с последействием подразделяются на потоки с простым и ограниченным последействием. Особенности потоков с простым последействием будут рассмотрены ниже.

Ординарность означает практическую невозможность группового поступления вызовов. Иначе говоря, вероятность поступления двух и более вызовов за любой бесконечно малый интервал времени Δt есть величина бесконечно малая более высокого порядка, чем Δt , то есть $p_{i \geq 2}(\Delta t) = o(\Delta t)$.

Для неординарных потоков вводится независимая от вызывающих моментов функция неординарности потока P_l , равная вероятности поступления группы из l вызовов в любой вызывающий момент.

Вызовы в инфокоммуникационных системах, как правило, ординарные из-за отсутствия синхронизма в инициации вызовов различными пользователями контента (А, V, D). Однако поток телеграмм к оператору для их обработки может быть неординарным, так как они могут поступать пачками.

К основным характеристикам случайных потоков вызовов относятся:

- ведущая функция;
- параметр;
- интенсивность.

Ведущая функция потока $\bar{x}(0, t)$ есть математическое ожидание числа вызовов в интервале $[0, t)$. Функция $\bar{x}(0, t)$ - неотрицательная, неубывающей и в практических задачах непрерывна и принимает только конечные значения.

Параметр потока $\lambda(t)$ в момент времени t есть предел отношения вероятности поступления не менее одного вызова в интервале $[t, t+\Delta t)$ к величине этого интервала при $\Delta t \rightarrow 0$:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{p_{i \geq 1}(t, t + \Delta t)}{\Delta t}.$$

Как видно, последняя формула полностью аналогична формуле для производной функции. Производная функции распределения – это плотность вероятности. Таким образом, параметр потока определяет плотность вероятности наступления вызывающих моментов в момент времени t .

Параметр стационарного потока λ есть математическое ожидание числа вызывающих моментов в единицу времени.

Интенсивность стационарного потока μ есть математическое ожидание числа вызовов в единицу времени.

Для нестационарных потоков вызовов различают среднюю и мгновенную интенсивности.

Средняя интенсивность $\mu(t_1, t_2)$ в интервале $[t_1, t_2)$ есть математическое ожидание числа вызовов в этом интервале в единицу времени:

$$\mu(t_1, t_2) = \frac{\bar{x}(0, t_2) - \bar{x}(0, t_1)}{t_2 - t_1}.$$

Мгновенная интенсивность – это производная ведущей функции потока по времени:

$$\mu(t) = \lim_{\Delta t \rightarrow 0} \frac{\bar{x}(0, t + \Delta t) - \bar{x}(0, t)}{\Delta t} = \bar{x}'(t).$$

Необходимо отметить, что интенсивность характеризует поток вызовов, а параметр характеризует поток вызывающих моментов. Поэтому всегда $\mu(t) \geq \lambda(t)$, а равенство имеет место лишь для ординарных потоков, когда в любой вызывающий момент поступает только один вызов. Аналогично для стационарных потоков имеем $\mu \geq \lambda$.

4.2. Простейший поток вызовов

Поток вызовов, обладающий тремя вышеприведёнными свойствами (стационарностью, ординарностью и отсутствием последействия), называется простейшим. Он описывается законом Пуассона:

$$P_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \quad (4.1)$$

где P_i – вероятность поступления за время t ровно i вызовов ($i = 0, \infty$), λ – параметр потока, равный среднему числу вызовов за единицу времени.

Расчёт вероятностей P_i можно проводить в следующей последовательности:

$P_0(t) = e^{-\lambda t}$, $P_1(t) = (\lambda t)e^{-\lambda t}$ и далее по рекуррентной формуле:

$$P_{i+1}(t) = P_i(t) \lambda t / (i+1)$$

Простейший поток вызовов (или пуассоновский поток) возникает в тех случаях, когда этот поток объединяет большое количество (теоретически бесконечное) небольших потоков, ни один из которых не является преобладающим. Как было показано выше простейший поток возникает в тех случаях, когда число источников вызовов $N \gg v$, где v – число обслуживающих устройств (каналов). Если $N > v$, то этот поток, как будет показано ниже, классифицируется как примитивный.

Наиболее типичным примером простейшего потока являются потоки вызовов в телефонных сетях. Например, телефонные сети общего пользования (ТфОП), речевой трафик в сотовых сетях, трафик IP-телефонии в Internet. В то же время компьютерный трафик носит явно выраженный пачечный характер, так как формируется потоком группы пакетов в момент появления на входе в сеть длинного (многопакетного) сообщения. Более того даже распределение моментов возникновения сообщений, как и моментов начала сессий, по многочисленным наблюдениям отличается от пуассоновского. Наиболее часто простейший поток описывается либо пуассоновским распределением числа вызовов потока λ за время t , либо экспоненциальным распределением интервалов между вызовами.

а) Варианты пуассоновского распределения (1) представлены на рис. 4.1 как зависимости вероятности $P_i(t)$ от числа вызовов (i), поступивших за время t для значений параметров: $\lambda t = 0,5$; $\lambda t = 3$; $\lambda t = 10$. Из графиков видно, что с увеличением λt огибающая смещается вправо и принимает симметричный вид приближаясь к нормальному закону распределения.

Для пуассоновского распределения существует очевидное равенство:

$$\sum_{i=0}^{\infty} P_i(\lambda t) = 1$$

Математическое ожидание числа вызовов для потока λ за время t равно:

$$M i = \sum_{i=1}^{\infty} i P_i(\lambda t) = \sum_{i=1}^{\infty} i \frac{(\lambda t)^i}{i!} e^{-\lambda t} = \lambda t$$

Характерной особенностью пуассоновского распределения является равенство математического ожидания и дисперсия, т. е. $D i = M i = \lambda t$. Поэтому при практических исследованиях для проверки гипотезы о принадлежности

какого-либо потока к пуассоновскому в первую очередь проверяется равенство $D_i = M_i$ и только потом делаются более детальные анализы.

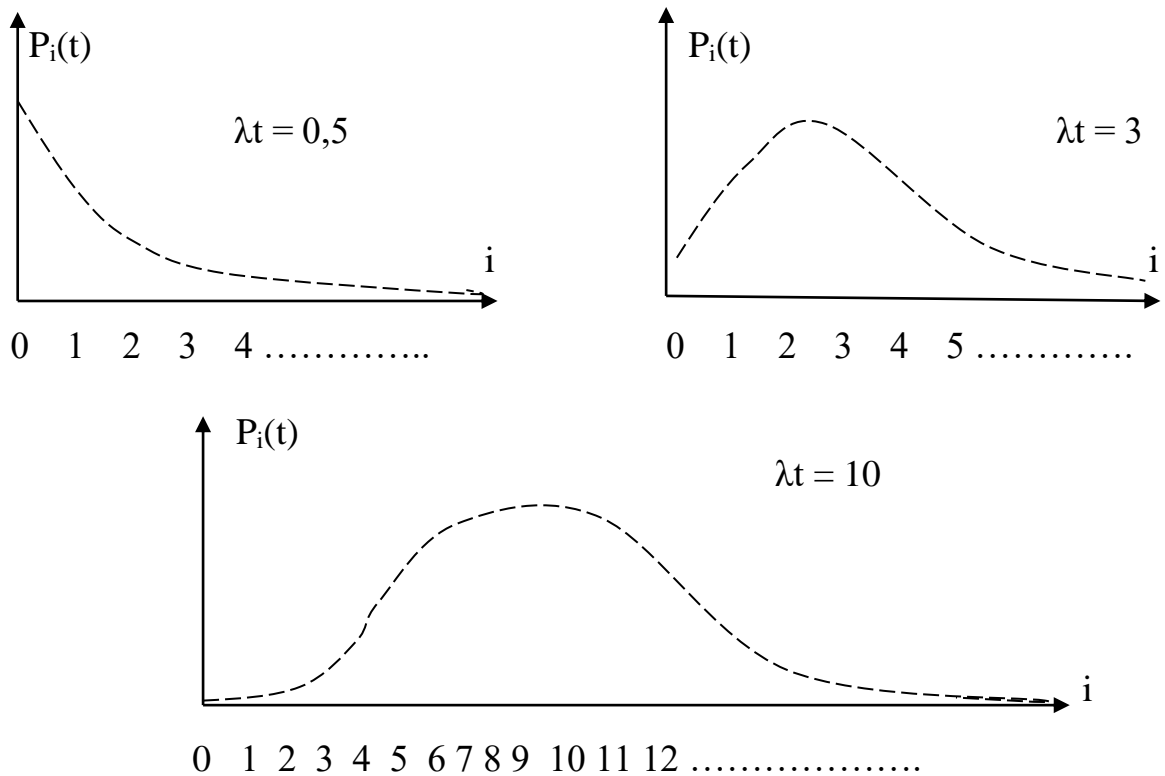


Рис. 4.1. Огибающие распределений пуассоновского потока

Интересные свойства пуассоновского потока, используемые при анализе потоков в сетях связи:

- сумма n пуассоновских потоков с параметрами λ_i тоже пуассоновский поток с параметром $\lambda = \sum_{i=1}^n \lambda_i$;

- разделение пуассоновского потока λ на n потоков с вероятностью P_i , где $\sum_{i=1}^n P_i = 1$ создаёт n пуассоновских потоков с параметрами $\lambda_i = \lambda P_i$.

б) Экспоненциальное распределение (рис. 4.2.) определяет вероятность $P(z < t)$ того, что от некоторого момента $t = 0$ интервал до очередного вызова (z), будет меньше переменной t . Преобразуем $P(z < t) = 1 - P(z > t)$. Вероятность $P(z > t)$ равносильна вероятности того, что внутри промежутка длиной t не поступит ни один вызов, а это в соответствии с (4.1) определяется как $P_0(t) = e^{-\lambda t}$. Таким образом, находим:

$$P(z < t) = 1 - P(z > t) = 1 - P_0(t) = 1 - e^{-\lambda t}. \quad (4.2)$$

Дифференцируя (4.2) по t , находим плотность экспоненциального распределения $p(t) = \lambda e^{-\lambda t}$ с параметром λ . Для экспоненциального распределения определены математическое ожидание длительности промежутков $Mz = 1/\lambda$, дисперсия $Dz = 1/\lambda^2$ и среднее квадратическое отклонение $\sigma z = \sqrt{Dz} = 1/\lambda$. При анализе случайных последовательностей нарушение равенства $Mz = \sigma z$ сразу

отвергает гипотезу об экспоненциальном распределении промежутков, а следовательно, и гипотезу о простейшем потоке.

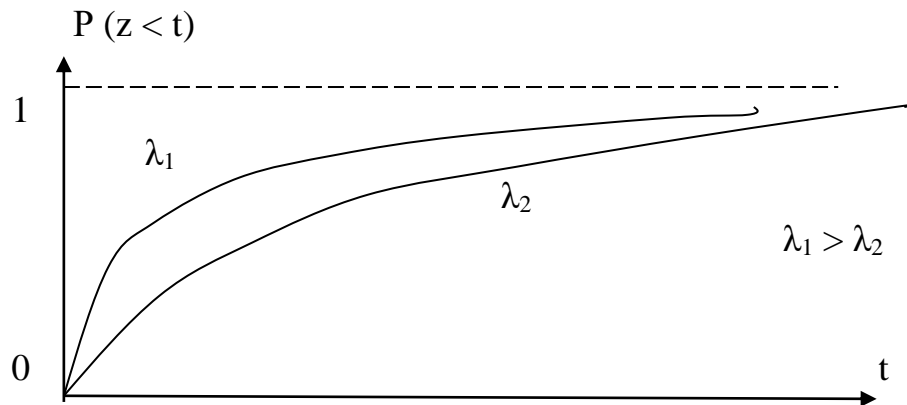


Рис. 4.2. Экспоненциальное распределение интервалов между вызовами

4.3. Потоки с простым последствием

Ординарный поток, параметр которого $\lambda_{r(t)}$ полностью определяется состоянием $r(t)$ обслуживающей системы в момент t , называется потоком с простым последствием. В данном случае последствие выражается косвенно через некоторые параметры, например, такие как количество абонентов, повторяющих свои вызовы. При простом последствии не отслеживаются моменты получения отказов различными абонентами. Этих моментов может быть очень много и в таких условиях невозможно получить обозримые аналитические выражения. Вместо этого рассматривается влияние на входной поток вызовов только общего числа этих абонентов. Именно в этом состоит смысл простого (в смысле упрощённого) последствия. Примерами потоков с простым последствием являются потоки с повторными вызовами и примитивные потоки.

а) Потоки с **повторными** вызовами возникают в системах с потерями в тех случаях, когда абонент, получив отказ в установлении соединения, не покидает систему, а делает повторные вызовы для передачи того же самого сообщения. В этом случае в систему поступает суммарный поток, состоящий из первичных и повторных вызовов:

$$\lambda_j = \lambda + j\beta.$$

Здесь λ – входной поток первичных вызовов, j – число источников, повторяющих вызовы, β – интенсивность потока повторных вызовов от одного абонента, которая связана с интервалами времени между повторами соотношением $\beta = 1/t_{\text{п}}$.

В данном случае индекс j показывает, что суммарный входной поток зависит от числа источников, повторяющих вызовы, т.е. косвенно от предыстории событий.

б) **Примитивным** называется ординарный поток с последствием, у которого параметр λ изменяется пропорционально числу свободных источников нагрузки:

Примером примитивного потока являются потоки вызовов в учрежденческой АТС, в которой новые вызовы могут инициироваться только свободными абонентами, т.е. абонентами, которые в данный момент не заняты разговором. А так как число абонентов УАТС ограничено, то переменным является и число свободных абонентов. Поэтому интенсивность входного потока равна:

$$\lambda_i = \alpha N_i = \alpha(N - i)$$

где α – параметр одного источника (интенсивность вызовов) в свободном состоянии, N – общее число источников вызовов, i – число обслуживаемых вызовов.

В данном случае индекс i показывает, что суммарный входной поток зависит от числа свободных источников, т.е. косвенно от предыстории событий

Вероятность поступления вызовов в примитивном потоке подчиняется распределению Бернулли:

$$p_k(t) = C_N^k a^k (1 - a)^{N-k},$$

где $a = \frac{\alpha}{1 + \alpha}$ – интенсивность нагрузки от одного источника.

Примитивный поток, как любой поток с потерями, тоже может инициировать повторные вызовы. В этом случае интенсивность входного потока будет равна:

$$\lambda_{ij} = \alpha(N - i - j) + j\beta.$$

Здесь число свободных источников уменьшается как за счёт разговаривающих абонентов (i), так и за счёт тех, которые повторяют свои вызовы (j).

4.4. Потоки мультисервисных вызовов

Характер потоков в современных сетях. Классическая теория телеграфика долгое время развивалась исключительно для нужд телефонии. Изучение поведения телефонного абонента привело к выводам о пуассоновском характере потока телефонных вызовов и экспоненциальном распределении продолжительности разговора.

Отличительной особенностью потока телефонных вызовов является его однородный характер. И хотя существует большое количество работ, посвящённых разно приоритетным вызовам, обслуживание этих вызовов производится, как правило, с одинаковой интенсивностью.

В эпоху мультисервисных сетей новым типичным абонентом становятся операторы персональных компьютеров (РС), хотя сетевые аналитики предпре-

кают в ближайшее время опережающий рост трафика межмашинного обмена (M2M).

Исследования поведения операторов показали:

- потоки пакетов, а в большинстве случаев и потоки требований на установление сеансов связи не являются пуассоновскими;
- длительности сеансов описываются распределениями не показательного типа, а логонормального, т.е. характеризуются наличием “весомых хвостов”;
- поступления пакетов, а также требования на установление сеансов связи коррелированы.

Главный недостаток пуассоновских моделей – недооценка образования пачечной нагрузки. Кроме того, обмен сообщениями в компьютерных сетях, как правило, носит эхоплексный характер (подтверждения о приёме, ответы на запросы в услугах клиент-серверных систем и др.).

Исследования различных типов сетевого трафика за последние полтора десятка лет доказывают, что сетевой трафик является не пуассоновским, а самоподобным (self-similar) или фрактальным (fractal) по своей природе. «Самоподобие» представляет собой свойство процесса сохранять свое поведение и внешние признаки при рассмотрении в разном масштабе. Из этого следует, что используемые методы моделирования и расчета сетевых систем, основанные на использовании пуассоновских потоков, не дают полной и точной картины происходящего в сети.

Кроме того, самоподобный трафик имеет особую структуру, сохраняющуюся при многократном масштабировании. В реализации, как правило, присутствует некоторое количество выбросов при относительно небольшом среднем уровне трафика. Данное явление ухудшает характеристики (увеличивает потери, задержки, джиттер пакетов) при прохождении самоподобного трафика через узлы сети. На практике это проявляется в том, что пакеты, при высокой скорости их движения по сети, поступают на узел не по отдельности, а целой пачкой, что может приводить к их потерям из-за ограниченности буфера, рассчитанного по классическим методикам.

На рис. 4.3 представлены временные реализации реального сетевого (самоподобного) трафика (слева) и традиционной несамоподобной (пуассоновской) модели телетрафика (справа) при различных масштабах временной оси. Сверху вниз масштаб временной оси укрупняется и кривая пуассоновской нагрузки всё больше приближается к своему среднему значению.

При реальных значениях вероятностей потерь (например, при $P < 0,1$) потери определяются именно выбросами нагрузки, а не её средними значениями. Поэтому расчёт канальной скорости (или числа каналов) при сглаженном характере нагрузки (пуассоновская нагрузка) приведёт к существенным ошибкам, т.е. реально будут возникать гораздо большие потери, чем требуемые для нормальной работы системы.

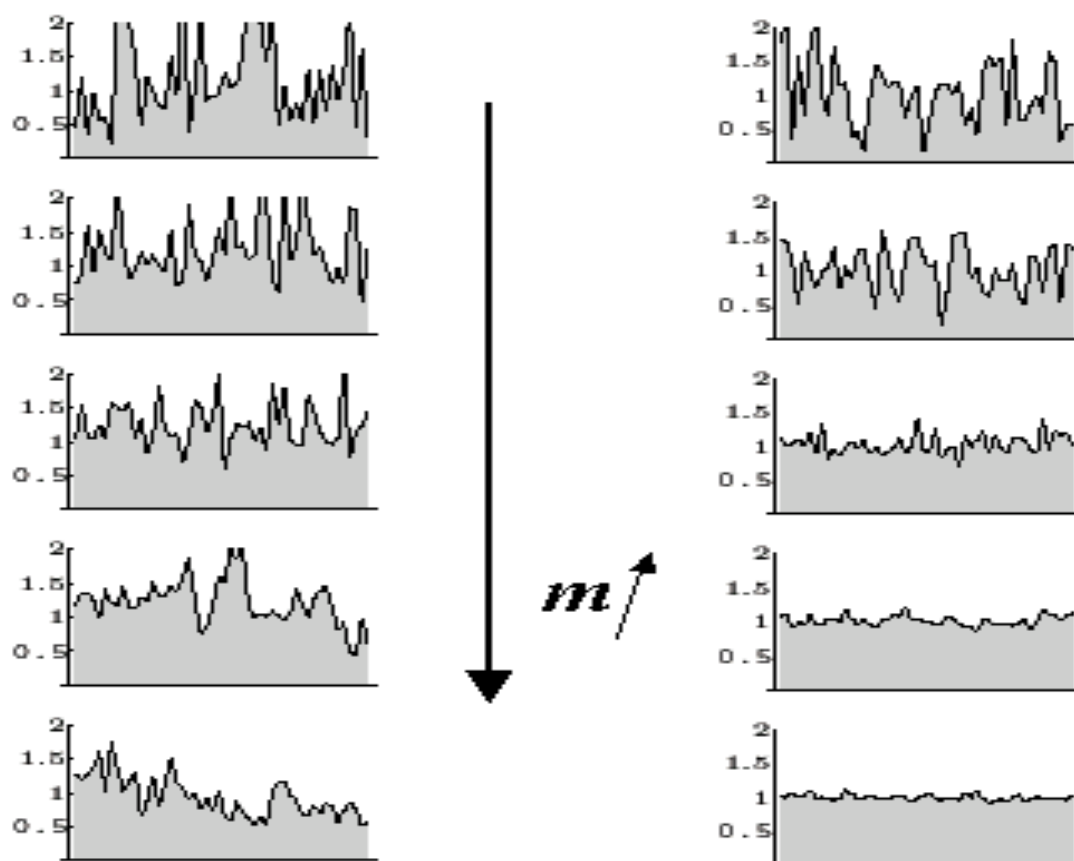


Рис. 4.3. Временные реализации самоподобного и несамоподобного трафиков

4.5. Расчёт мультисервисных потоков.

Главной особенностью потоков в современных сетях является переход от требований на предоставление стандартного телефонного канала в режиме коммутации каналов к формированию потока пакетов для любых видов трафика и обслуживанию этого потока в режиме коммутации пакетов. В сетях Internet принято определять потоки нагрузок как некоторый контент Triple Play (A, V, D). При этом, если аудио и видео информация носит явно выраженный потоковый характер (интенсивность входного потока пакетов зависит только от создающего его приложения и используемой процедуры сжатия), то трафик данных по своей природе является всплесковым, так как возникает в момент инициализации обмена, и его интенсивность в какой-либо точке пути определяется не характером данных, а пропускной способностью этого пути.

В связи с этим, в самом общем случае можно говорить о потоке пакетов на входе сети или какого-либо узла сети (например, порт коммутатора или маршрутизатора) как о сумме отдельных потоков

$$\Lambda = \sum_{i=1}^k \lambda_i \text{ [пакеты/ед. вр.]},$$

где i – вид входного потока, $i = 1, k$.

Можно, конечно, сказать, что, если составляющих потоков λ_i достаточно много и среди них нет превалирующего, то суммарный поток Λ будет простейшим. Но такое объединение ни коим образом не облегчает задачи анализа, так как обслуживание этих потоков в сети должно вестись по существенно разным алгоритмам.

Элементы мультисервисных сетей (каналы, порты и др.) обладают конечной пропускной способностью, измеряемой обычно в битах. Поэтому желательно для любой комбинации потоков иметь соответствующий битовый эквивалент – битовый входной поток – C [бит/с], чтобы сопоставлять возможности элементов сети с поступающим потоком. Такой битовый поток можно определить как

$$C = 8 \sum_{i=1}^k \lambda_i Q_i = \sum_{i=1}^k \lambda_i q_i$$

где λ_i – интенсивность входного потока пакетов i -го вида в секунду,

Q_i – длина пакета в байтах, q_i – длина пакета в битах.

В простейшем случае, если на элемент сети поступают три битовых потока: аудио – C^A , видео C^V и данных C^D , мы получим суммарный битовый поток:

$$C = C^A + C^V + C^D. \quad (4.3)$$

Потоки в сетях. Сложность анализа потоков в сетевых структурах состоит в том, что многочисленные и, как правило, разнородные потоки используют одни и те же ресурсы и конкурируют за использование этих ресурсов. Потоки в сетях, а конкретнее – информационное тяготение между узлами сети задаётся обычно в виде матрицы потоков (матрицы тяготений) размерности n , где n – число узлов в сети. Элементами этих матриц являются потоки между соответствующими узлами, выраженные в пакетах, байтах или битах. Как правило, интенсивность информационного обмена выражается в пакетах (чаще в байтах – как более универсальной мере), а потоки, предназначенные для сопоставления с пропускными способностями каналов – в битах. При этом для определения суммарных потоков от составляющих разных категорий используется обычное правило сложения матриц. Так, например, по аналогии с (4.3) можно найти суммарную матрицу битовых потоков для контента Triple Play.

$$C = \|c_{ij}\| = \|c_{ij}^A\| + \|c_{ij}^V\| + \|c_{ij}^D\| \quad (4.4)$$

5. Характеристики качества обслуживания

МСЭ рассматривает качество обслуживания как результат совместного проявления характеристик обслуживания. Этот результат определяет степень удовлетворенности пользователя предоставленной ему услугой.

Степень удовлетворенности уровнем обслуживания может оцениваться такими характеристиками:

- обеспечение обслуживания (service support);
- управляемость обслуживания (service operability);
- возможности обслуживания (serveability);
- безопасность обслуживания (service security).

Характеристики **обеспечения обслуживания** отражают способность Оператора предоставлять услуги и способствовать их использованию. Характеристики **управляемости** обслуживания оценивают удобство и простоту пользования услугами. **Возможности** обслуживания, в свою очередь, делятся на три группы, для которых в рекомендации МСЭ E.800 предлагаются такие характеристики:

- доступность услуг (service accessibility);
- непрерывность обслуживания (service retainability);
- полноценность обслуживания (service integrity).

Характеристики доступности услуг оценивают возможность их получения (с заранее специфицированными допусками и с соблюдением других заданных условий) по запросу пользователя. Характеристики непрерывности обслуживания определяют возможность пользования полученной услугой с заданными атрибутами в течение запрошенного интервала времени. Характеристики полноценности обслуживания – меры того, что обслуживание, будучи полученным, происходит без значительного ухудшения.

Характеристики безопасности связаны с такими аспектами функционирования сети связи, как несанкционированный мониторинг, жульническое использование, злонамеренное повреждение, неправильное применение, ошибка человека, стихийное бедствие.

Мультисервисная сеть обслуживает все виды трафика. Однако, устанавливать одинаковые требования к показателям качества обслуживания (QoS) для всех видов трафика не представляется разумным по техническим и экономическим соображениям. МСЭ в рекомендации Y.1541 выделил шесть классов (COS – Class of Service), различающихся величинами показателей QoS. В таблице 5.1 приведены значения трех показателей QoS для шести классов. Эти значения определяются для следующих показателей:

- IPTD – задержка переноса IP пакетов,
- IPDV – вариация задержки IP пакетов,
- IPLR – доля потерянных IP пакетов.

Значения доли искаженных IP пакетов (IREP) в таблице 5.1 не приводятся, этот показатель не имеет отношения к рассматриваемым в данном контексте вопросам. Символ "U" (первая буква в слове "unspecified") указывает на то, что показатель для данного класса обслуживания не нормируется.

Таблица 5.1. Классы обслуживания для IP пакетов

Класс QoS	IPTD	IPDV	IPLR
0	100 мс	50 мс	10^{-3} .
1	400 мс	50 мс	10^{-3} .
2	100 мс	U	10^{-3} .
3	400 мс	U	10^{-3} .
4	1 с	U	10^{-3} .
5	U	U	U

Примечания:

1) При большом времени распространения сигналов могут возникать сложности для классов "0" и "2" с соблюдением норм на среднее значение времени задержки IP пакетов. Величины IPTD определены для максимальной длины информационного поля пакета 1500 байтов.

2) Величина IPDV определяется разницей между верхней границей, в качестве которой рекомендуется 99,9% квантиль, и нижней границей задержки, измеренной в течение интервала оценки. В качестве длительности этого интервала предлагается выбирать одну минуту. Все эти соображения МСЭ считает предварительными и требующими дополнительного изучения.

3) Эта величина зависит от емкости тракта обмена пакетами. Приемлемая величина вариации достигается для трактов с пропускной способностью 2048 кбит/с и более, а также при длине информационного поля пакетов менее 1500 байтов.

Класс обслуживания "0" предназначен для обмена информацией в реальном времени (в частности, для речи с использованием технологии VoIP). Он предусматривает создание отдельной очереди с приоритетной обработкой пакетов. Для класса обслуживания "0" характерны ограничения на принципы маршрутизации и допустимое расстояние между взаимодействующими терминалами (время распространения сигналов). Интерактивность для класса "0" определяется как "высокая" – high.

Класс обслуживания "1" также предназначен для обмена информацией в реальном времени, но с менее жесткими требованиями. Поэтому накладываются менее жесткие ограничения на принципы маршрутизации и время распространения сигналов, чем для класса "0". Также предусматривается создание отдельной очереди с приоритетной обработкой пакетов.

Класс обслуживания "2" ориентирован на обмен данными с высокой степенью интерактивности и ему, как и классу "0", присвоен уровень высокой интерактивности. К этому классу относится и сигнальная информация. Для класса обслуживания "2" характерны такие же ограничения на принципы маршрутизации и время распространения сигналов, как для класса "0". Для пакетов этого класса формируется своя очередь на обработку, которая осуществляется со вторым приоритетом. Это означает, что пакеты классов "0" и "1" имеют преимущество на обработку.

Классу обслуживания "3", предназначенному для обмена с менее высоким уровнем интерактивности, присущи те же ограничения на принципы маршрутизации и время распространения сигналов, что и классу "1". Обслуживание пакетов этого класса должно осуществляться со вторым приоритетом.

Класс обслуживания "4" предназначен для обмена различной информацией с низкой вероятностью потери (короткие транзакции, потоковое видео и прочие). Допускаются длинные очереди пакетов на обработку, которая осуществляется со вторым приоритетом. Никакие ограничения на маршрутизацию и время доставки сообщений не накладываются.

Класс обслуживания "5" ориентирован на те IP-приложения, которые не требуют высоких показателей QoS. Соответствующие пакеты формируют отдельную очередь; обслуживание осуществляется с самым низким приоритетом (в данном случае он имеет третий номер). Никакие ограничения на маршрутизацию и время доставки сообщений не накладываются.

Сама форма таблицы 5.1 говорит о том, что заявки в мультисервисной сети обслуживаются по алгоритму с приоритетами и ожиданием. Для незначительной доли заявок (вероятность такого события составляет 0,001) допускаются потери.

6. Дисциплины обслуживания

Наиболее употребимые в современных ТК-структурах дисциплины обслуживания вызовов представлены на блок-схеме рис. 6.1.



Рис. 6.1. Дисциплины обслуживания

Дисциплина «Без потерь» предусматривает обслуживание поступившего вызова в любом случае и без какой-либо задержки. Для простейшего потока вызовов (теоретически бесконечное число источников) это означает наличие бесконечного числа обслуживающих устройств (например, каналов). При примитивном потоке, характерном для учреждений АТС с ограниченным

числом источников вызовов N , обслуживание без потерь возможно только, если число источников не превышает число каналов v , т.е. если $N \leq v$. Практически такое соотношение встречается редко, как экономически нецелесообразное. Поэтому обслуживание с дисциплиной без потерь носит, в основном теоретический характер.

Обслуживание с явными потерями предполагает, что вызов и соответствующее ему сообщение при получении отказа в немедленном обслуживании полностью теряются и на обслуживание больше не поступают, по крайней мере, в данную систему. Обычно наблюдаемое желание абонента повторить свой вызов для передачи того же самого сообщения соответствует дисциплине с повторными вызовами.

Условные потери выражаются в задержке передачи сообщения сверх допустимого срока. Понятие – допустимый срок – уточняется в каждом конкретном случае и вводится в связи с тем, что для любой коммутационной системы, а тем более для сети необходимо не нулевое время хотя бы для организации обслуживания нового вызова.

Обслуживание с условными потерями может быть организовано по системе с ожиданием соединения и с повторением вызовов. В первом случае задержанные источники становятся в очередь и обслуживаются по мере освобождения обслуживаемых устройств. Во втором случае, не обслуженные источники образуют как бы активную очередь и повторяют вызовы через случайные или детерминированные промежутки времени вплоть до установления соединения. При этом разграничиваются понятия потери вызовов и потери сообщений, поскольку потеря отдельных вызовов (попыток) не ведёт к явной потере сообщения. Происходит только задержка обслуживания, т.е. возникают условные потери (потери времени) как и в системе с ожиданием.

Дисциплина обслуживания с повторением вызовов математически более общая модель по сравнению с дисциплинами с явными потерями и с ожиданием. С ростом интенсивности повторения промежутки между соседними попытками уменьшаются и при их стремлении к нулю все попытки сливаются в непрерывное ожидание. С другой стороны, с уменьшением интенсивности повторения, когда промежутки между соседними попытками могут увеличиться настолько, что повторные вызовы будут как бы теряться среди первичных, данная модель приближается к модели с потерями. Так как математическая модель системы с потерями проще, чем системы с повторными вызовами, то на практике при малых потерях вызовов ($P \leq 0,01$) применяются модели с потерями.

Наряду с основными дисциплинами обслуживания – с явными и условными потерями – возможен ряд комбинированных дисциплин. В системах с ожиданием часто ограничиваются время пребывания в очереди или длина очереди, или оба ограничения вместе. При обслуживании с повторением возможны ограничения на число повторных попыток или число источников, повторяющих вызовы (условная длина очереди).

Модели с комбинированными потерями позволяют определить более реальные – с практической точки зрения – алгоритмы обслуживания заявок. Обслуживание с ограничением процедурного времени давно используется в телефонных станциях. Например, если Вы слышите акустический сигнал "Ответ станции", но не набираете номер в течение некоторого интервала времени, то обслуживание будет прервано. Вы услышите акустический сигнал "Занято".

С алгоритмом, который ограничивает число мест для ожидания, многие абоненты сталкиваются при попытке дозвониться до справочной службы "09". Если все операторы заняты, Вас поставят в очередь и попросят подождать. Если же заняты и все места для ожидания (Заполнен буфер ожидающих запросов), то вежливый голос приносит Вам свои извинения и просит повторить вызов позже (потеря вызова). Некоторые справочные системы, используемые в телефонной сети, сочетают оба вида ограничений – по длительности ожидания и по числу мест в очереди.

Заявки могут обрабатываться без приоритета. Такой алгоритм характерен, например, для устройств управления электромеханических коммутационных станций. Приоритетные стратегии обработки заявок можно разделить на три основные группы, которые следует рассмотреть подробнее. Комбинированные алгоритмы предусматривают переход к приоритетному обслуживанию при определенных условиях (например, резкий рост трафика, приводящий к снижению показателей качества обслуживания).

Для анализа приоритетных стратегий целесообразно ввести простую модель. Все заявки, поступающие в систему, делятся на группы, которым присваивается приоритет от 1 до k . Заявка с меньшим номером приоритета имеет преимущество перед заявками с большим номером.

Приоритеты могут быть относительные и абсолютные. В системе с относительными приоритетами обслуживание заявок не прерывается. Допустим, заявка с приоритетом j застала все обслуживающие приборы занятыми. Тогда она встает в очередь перед всеми заявками, имеющими меньший приоритет, т.е. для которых $i > j$.

Системы с абсолютными приоритетами основаны на прерывании обслуживания заявок. Такая возможность предусмотрена для всех случаев, когда обслуживаются заявки более низкого приоритета. При этом могут использоваться различные варианты возобновления прерванного процесса обслуживания заявок. Необходимость в абсолютных приоритетах может возникать, например, в случаях, когда очень срочный управляющий пакет не может ждать окончания передачи длинного и не очень срочного пакета (например, в технологии Frame Relay длина пакета может достигать до 4500 байт).

Более 100 лет – время почти исключительного господства телефонной связи – основной дисциплиной являлось обслуживание с потерями. По такому принципу функционируют и современные сети с коммутацией речевых каналов в ТфОП и в сотовых сетях (поколения 2G и 3G). В то же время системы с коммутацией пакетов с любым видом трафика (A, V, D) реализуют

дисциплину с ожиданием, поскольку пакет, поступающий в узел коммутации сети (например, в маршрутизатор), при занятости канала нужного направления, помещается в очередь (буфер ожидания) и будет вводиться в канал по мере его освобождения с соблюдением установленных правил по обслуживанию разноприоритетных видов трафика. По такому принципу функционируют Internet и пакетные режимы в сотовых сетях (GPRS, EDGE и сотовая сеть четвёртого поколения – LTE).

Строго говоря, в пакетных сетях применяется комбинированная дисциплина обслуживания, так как буферы ожидания для пакетов всегда ограничены по объёму, а контроль времени пребывания пакетов в любом элементе сети и в сети в целом является обязательным требованием для сетевых алгоритмов.

7. Нагрузка, работа и их виды

7.1. Общие положения

Для характеристики процесса обслуживания в сетях и отдельных устройствах используются понятия нагрузки и работы. Нагрузка, обслуживаемая многоканальным устройством в момент времени t , есть число занятых линий $i(t)$ в этот момент. При измерениях на конкретном пучке определяют среднюю интенсивность обслуженной нагрузки как среднее число занятых линий за определённый промежуток времени. Интенсивность нагрузки измеряется в эрлангах (Эрл). Одна непрерывно занятая линия обслуживает нагрузку в 1 Эрл.

Образование многоканальных пучков является спецификой потоковых видов трафика (A, V), при которых каждый канал монопольно закрепляется за конкретным информационным обменом (телефонный разговор, телепрограмма, видеоконференция). При этом скоростной канал, например, E1 (2048 кбит/с) разбивается на 30 каналов ОЦК (64 кбит/с).

В компьютерных сетях с трафиком всплескового характера не целесообразно производить такое разделение. Можно показать, что при разделении, например, одного канала 1GB-Ethernet на 10 каналов FastEthernet, т.е. если разбить скорость 1000 Мбит/с на 10 каналов по 100 Мбит/с, то время передачи пакетов по таким каналам увеличится.

В этих случаях нагрузку, обслуженную одним каналом, можно оценивать только в долях эрланга, измеренную как доля времени за определённый промежуток, в течении которого канал был занят. При этом обслуженная нагрузка $0 \leq Y_o \leq 1$ Эрл.

Работу по передаче информации, выполненную системой обслуживания за промежуток $[t_1, t_2]$, можно определить как суммарное время занятия всех линий за этот промежуток, или как произведение средней интенсивности обслуженной нагрузки $Y_o(t_1, t_2)$ на величину этого промежутка, т.е.:

$$A = Y_o(t_1, t_2) * (t_2 - t_1)$$

Работа измеряется в часо-занятиях или в Эрл-часах (пишется: ч-зан или Эрл-ч).

Наряду с обслуженной нагрузкой, важное значение при анализе систем имеет поступающая нагрузка, которая определяется как произведение интенсивности потока вызовов на среднее время одного занятия: $Y = \lambda t$. Разность между поступающей и обслуженной нагрузками называется избыточной нагрузкой – $Y_{и}$. Все три вида нагрузки измеряются в Эрлангах. При этом поступающая нагрузка $Y = Y_o + Y_{и}$.

Работа, как и интенсивность нагрузки, обладает свойством аддитивности.

Нагрузка идеальной системы без потерь называется потенциальной.

Разность между потенциальной и обслуженной нагрузками называется потерянной нагрузкой.

Состояние любого обслуживающего прибора может быть представлено с помощью периодов занятости и свободности. За исключением систем, работающих по расписанию, длительность каждого из этих двух периодов представляет собой случайную величину.

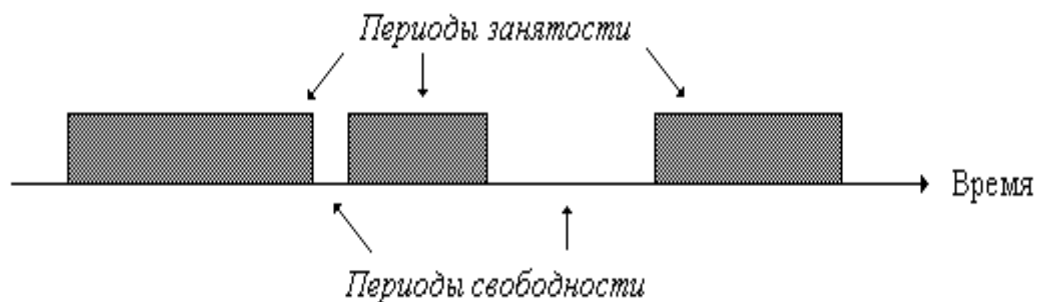


Рис. 7.1. Два состояния обслуживающего прибора

Для телефонной нагрузки весьма существенны три параметра. Во-первых, важно знать число источников нагрузки – N . Характерные примеры этого параметра – число абонентов, подключённых к АТС, или число терминалов, включенных в концентратор. Во-вторых, необходимо определить среднее число вызовов, поступающих от одного источника за единицу времени – λ . В третьих, следует оценить математическое ожидание длительности обслуживания одного вызова – t . Тогда поступающая нагрузка определится как

$$Y = N\lambda t$$

7.2. Изменение нагрузки во времени

Нагрузка в телекоммуникационных системах заметно меняется в течение дня (колебания можно проследить также по дням недели, месяцам, годам). Расчет оборудования, необходимого для обслуживания, например, телефонного трафика, осуществляется для часа наибольшей нагрузки (ЧНН). В некоторых публикациях предлагается – в дополнение к ЧНН – выделять пе-

риод наибольшей нагрузки. Его длительность ($T_{\text{МАХ}}$) может быть различной в зависимости от цели, для которой необходимо знать период максимальной нагрузки. Она может быть меньше или больше одного часа. В рассматриваемом случае (Рис. 7.2) $T_{\text{МАХ}}$ составляет три часа.

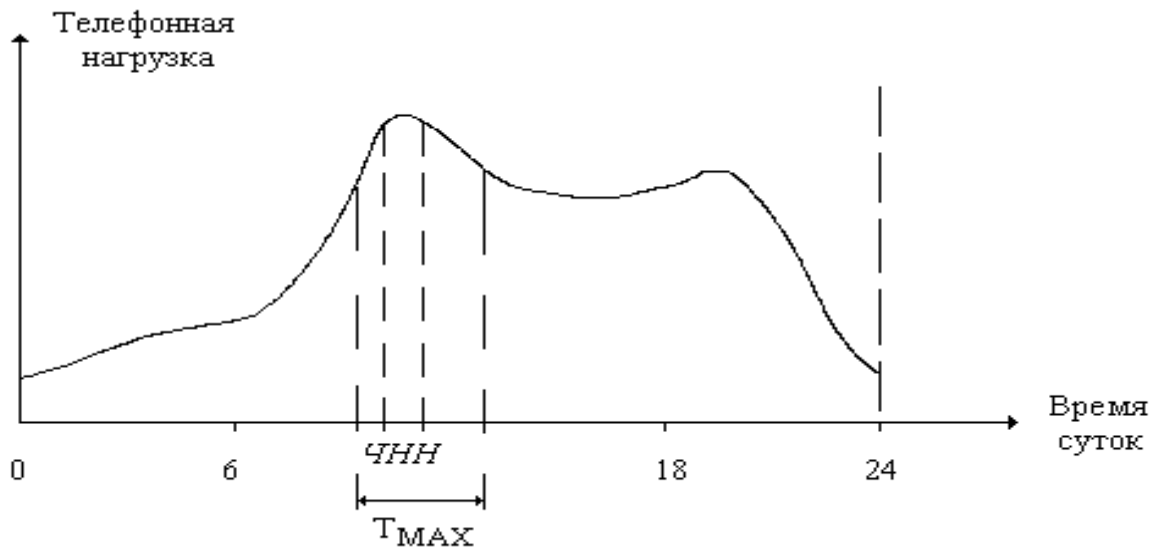


Рис. 7.2. Час и период наибольшей нагрузки

Доля нагрузки, создаваемой в ЧНН — $Y_{\text{ЧНН}}$, часто оценивается коэффициентом концентрации — $k_{\text{ЧНН}}$. Для его оценки необходимо также знать среднесуточную нагрузку $Y_{\text{СУТ}}$. Тогда

$$k_{\text{ЧНН}} = \frac{Y_{\text{ЧНН}}}{Y_{\text{СУТ}}}.$$

Диапазон изменения этого коэффициента значителен. Для многих телефонных сетей в крупных городах он составляет около 0,1. В менее крупных сетях величина $k_{\text{ЧНН}}$ может достигать 0,2.

Выделение ЧНН или периода максимальной нагрузки преследует конкретную цель. Установленные Администрацией связи показатели качества обслуживания телефонной нагрузки должны соблюдаться и в периоды высокого трафика. Естественно, что в другое время (в частности, ночью) ресурсы телефонной сети будут использоваться не так эффективно как в ЧНН. Правда, Оператор с помощью технических и маркетинговых операций может обеспечить экономичное функционирование инфокоммуникационной системы. Например, можно предлагать пользователям льготный тариф для доступа в Интернет в ночное время.

8. Системы с явными потерями

8.1. Символическое обозначение исследуемых моделей

Для удобства дальнейшего изложения введём символику Кендалла – Башарина для систем массового обслуживания, которая была предложена более пятидесяти лет назад Кендаллом. Она была основана всего на трех символах. Для описания более сложных процессов функционирования современных инфокоммуникационных систем эта классификация была дополнена Башариным до шести символов.

Первый разряд символики определяет поступающий поток вызовов, второй – закон распределения времени обслуживания. Так символами M, D, E, G в первой позиции обозначают соответственно: экспоненциальное, детерминированное, эрланговское и произвольное (General) распределение промежутков между вызовами. Эти же символы во второй позиции означают соответствующие распределения длительностей обслуживания.

Символом v в третьей позиции указывается число обслуживающих приборов (каналов). В четвертой позиции символом r указывается число мест ожидания в очереди. Для систем с потерями этот символ опускается.

Например, символы M/G/7/5 обозначают, что рассматривается 7-ми канальная система, на которую поступает пуассоновский поток вызовов, обслуживание с произвольным распределением длительностей, число мест ожидания в очереди равно 5. Символами D/E/15 обозначают 15-и канальную систему с потерями, с детерминированным входным потоком и эрланговским распределением длительностей обслуживания, а сочетание G/D/1/ ∞ обозначает одноканальную систему с произвольным входным потоком, постоянной длительностью обслуживания и бесконечной очередью.

В специальной литературе применяется более расширенная и разнообразная символика.

8.2. Обобщенная модель и качество обслуживания

Рассмотрим v -канальную систему обслуживания M/M/ v (пуассоновский входной поток и экспоненциальная длительность обслуживания). Время обслуживания одного вызова примем за условную единицу времени $h = 1$ у.е.в. Число занятых каналов i ($i = 0, v$) назовём состоянием исследуемой системы. Параметр потока вызовов Λ_i выразим в числе вызовов/у.е.в. Его можно трактовать как интенсивность поступающей нагрузки в состоянии системы i . Смысл условной единицы времени в том, что с её помощью можно в дальнейших исследованиях абстрагироваться от конкретных единиц астрономического времени (секунды, часы) и воспользоваться тем фактом, что одна и та же поступающая нагрузка может быть выражена как $U = \lambda t$, так и как $U = \Lambda h$. А так как $h = 1$ у.е.в., то получается, что поступающая нагрузка U и поступающий в у.е.в. поток Λ численно равны. Запись $U = \Lambda$ в строгом смысле – не корректна, так как U измеряется в Эрлангах, а Λ в выз/у.е.в.

Качество обслуживания в системах с потерями оценивается вероятностями потерь по времени, по вызовам и по нагрузке.

Потери по времени – это доля времени, в течении которого заняты все каналы (приборы), т.е. $P_t = t_{зан}/T_{наб}$, где $t_{зан}$ – время занятости каналов, а сле-

довательно время, в течении которого новые вызовы получают отказ, а $T_{\text{наб}}$ – общее время наблюдения.

Потери по вызовам – это доля потерянных вызовов, т.е. $P_v = c_{\text{пот}} / C_{\text{общ}}$, где $c_{\text{пот}}$ – число потерянных вызовов, $C_{\text{общ}}$ – общее число поступивших вызовов.

Потери по нагрузке – это доля избыточной нагрузки, т.е. $P_n = Y_n / Y$, где Y_n – избыточная нагрузка, а Y – общая поступающая нагрузка. Напомним, что $Y = Y_o + Y_n$, где Y_o – обслуженная нагрузка.

Между этими оценками, в общем случае, существует следующее соотношение:

$$P_t \geq P_v \geq P_n$$

Равенство наступает для системы $M/M/v$, т.е. для простейшего входящего потока при экспоненциальной длительности обслуживания. В этом случае:

$$P_t = P_v = P_n$$

8.3. Частные случаи распределения вероятностей состояния

В таблице 8.1 представлены четыре частных случая обслуживающих систем:

2 вида обслуживания (без потерь и с потерями) и 2 типа входного потока (простейший и примитивный). Представленные формулы позволяют определить вероятность занятости в произвольный момент времени ровно i каналов или долю времени (на бесконечном интервале), когда занято i каналов.

Таблица 8.1. Частные случаи распределения вероятностей

Простейший входной поток		Примитивный входной поток	
Без потерь	С потерями	Без потерь	С потерями
Распределение Пуассона	Распределение Эрланга	Распределение Бернулли	Распределение Энгсета
$P_i = \frac{\Lambda^i}{i!} e^{-\Lambda}$ $i = 0, \infty$	$P_i = \frac{\frac{\Lambda^i}{i!}}{\sum_{j=0}^v \frac{\Lambda^j}{j!}}$ $i = 0, v$	$P_i = C_N^i a^i (1-a)^{N-i}$ $i = 0, N; \text{ при } v \geq N$	$P_i = \frac{c_N^i \alpha^i}{\sum_{j=0}^v c_N^j \alpha^j}$ $i = 0, v; \text{ при } v < N$

На рис. 8.1 представлено распределение Пуассона для числа занятых линий при $\Lambda = 3$ Эрл. Оно в точности совпадает с представленным на рис. 3.1 пуассоновским распределением числа вызовов на интервале t при $\lambda = 3$ выз./ед.вр. Оба распределения и в формульном виде и в графическом идентичны, но совершенно различны по смыслу.

Распределение Эрланга (усечённое распределение Пуассона) отражает наиболее реальную ситуацию при ограниченном числе каналов ($v < \infty$).

Наиболее важным в распределении Эрланга является состояние $i = v$, т.е. состояние, когда заняты все каналы. Разумеется, что это состояние, когда наступают потери (отказы в обслуживании), т.е.

$$P_i = P_v = P_t = P_B = P_H = E_v(\Lambda).$$

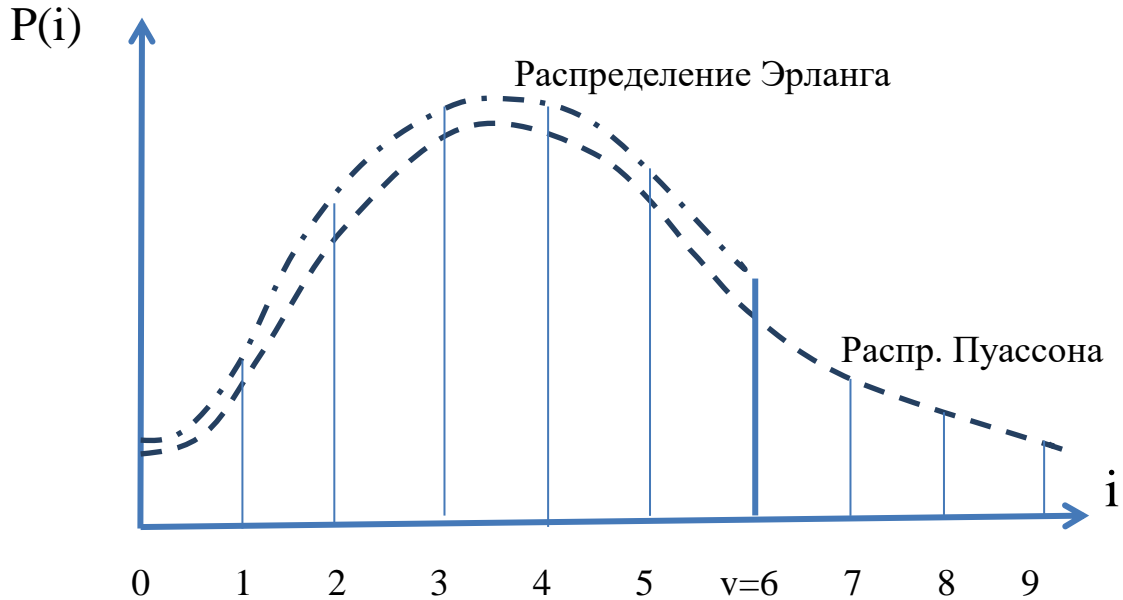


Рис. 8.1. Огибающие распределений вероятностей чисел занятых каналов для простейшего входящего потока.

Последнее обозначение $E_v(\Lambda)$ является общепринятым обозначением формулы Эрланга. Наиболее частая практическая задача определения необходимого числа линий в пучке для обслуживания поступающей нагрузки при заданной вероятности потерь данной формулой в явном виде не решается, т.к. невозможно выразить число каналов v как функцию от Λ и P . В связи с этими трудностями формула Эрланга табулирована в различных справочниках и задачниках. Кроме того в последнее время появилось несколько удобных программ, позволяющих очень легко из трёх величин – нагрузка (Λ), число каналов (v) и вероятность потерь P – найти любую при двух других известных.

Распределениям Бернулли и Энгсета подчиняются вероятности числа занятых линий при примитивном входном потоке, определяемом по соотношению:

$$\Lambda_i = \alpha N_i = \alpha(N - i)$$

где α – интенсивность одного источника в свободном состоянии (выз/у.е.в.),

N – общее число источников, N_i – число свободных источников, i – как везде в данном разделе – число занятых источников.

В формуле Бернулли величина a определяет вероятность одного успешного испытания и равна $a = \alpha / (1 + \alpha)$. Распределение Энгсета (или усечённое распределение Бернулли) описывает систему с потерями, так как $v < N$, а распределение Бернулли – систему без потерь ($v \geq N$).

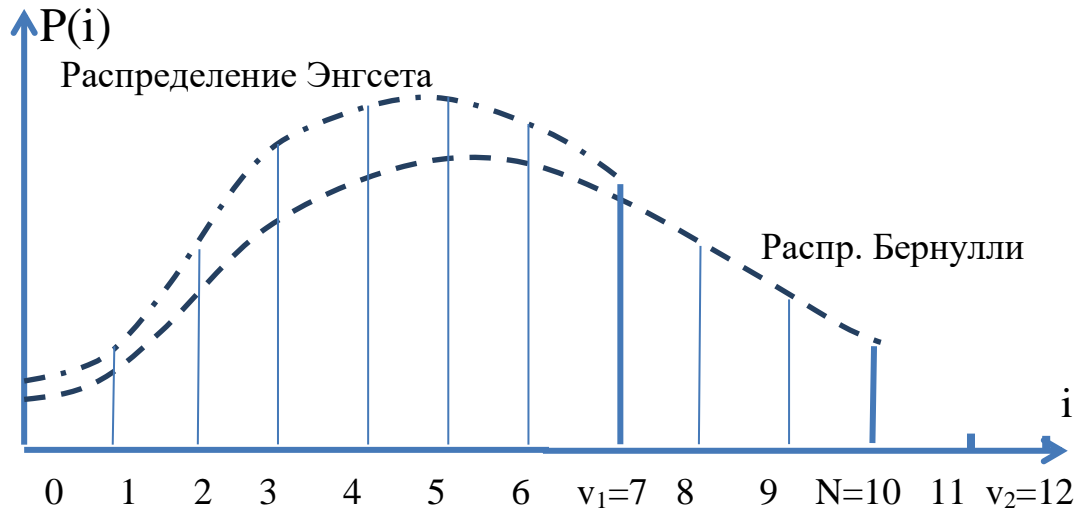


Рис. 8.2. Огибающие распределений вероятностей чисел занятых каналов для примитивного входящего потока.

8.4. Пропускная способность каналов для системы с потерями

Процедура занятия одного из свободных каналов в пучке может быть упорядоченной или случайной. При упорядоченной процедуре, как правило, занимается свободный канал с меньшим номером. При случайной – выбирается любой свободный канал.

В зависимости от реализованной процедуры занятия по разному определяется коэффициент использования канала η , под которым понимается доля времени занятости канала.

При случайном занятии вся обслуженная нагрузка равномерно распределяется между всеми каналами, поэтому использование каналов одинаково и равно:

$$\eta_{\text{сл}} = Y_o / v.$$

Отметим, что показатель η имеет тройной смысл:

- коэффициент использования канала;
- доля времени занятости канала;
- интенсивность нагрузки, обслуженной данным каналом в Эрлангах.

При упорядоченном занятии коэффициент использования i -го канала определяется как разность между избыточными нагрузками предыдущих каналов (фактически той нагрузки, которая является поступающей в i -ый канал) и данного канала (т.е. нагрузки, которую не смогли обслужить первые i каналов):

$$\eta_i = R_{i-1} - R_i = \Lambda E_{i-1}(\Lambda) - \Lambda E_i(\Lambda) = \Lambda[E_{i-1}(\Lambda) - E_i(\Lambda)]$$

Коэффициент использования последнего канала в пучке является очень важным показателем, так как именно он определяет вероятность потерь в пучке P_v , и может достигать единиц и даже долей процента. В то же время, каналы с меньшими номерами, которые занимают в первую очередь, обеспечивают основной пропуск трафика. В мощных пучках коэффициент использования первых каналов доходит до 99% и даже выше. При этом этот коэффициент тем выше, чем больше каналов в пучке для одинаковой вероятности потерь.

В связи с этим при проектировании сетей стараются организовывать как можно более мощные пучки, объединяя мелкие потоки с помощью коммутационных устройств. В целом такие решения позволяют экономить на стоимости каналов, правда за счёт затрат на коммутационные станции.

9. Системы с ожиданием

9.1. Системы с бесконечной очередью.

Рассматривается математическая модель $M/M/v/\infty$, т.е. v -канальная система, простейший входной поток, экспоненциальное обслуживание и возможность очереди бесконечной длины. За параметр входного потока, как и в системе с потерями, принимается величина Λ выз./у.е.в., которая численно равна поступающей нагрузке. Напомним, что $Y = \Lambda h = \lambda t$, а $h = 1 \text{ у.е.в.} = t_{\text{обсл}}$

При занятости всех v каналов поступивший вызов становится в очередь и обслуживается после некоторого ожидания. В отличие от системы с потерями, где за состояние системы i принималось число занятых каналов, в системах с ожиданием под состоянием системы понимается суммарное количество вызовов, находящихся в очереди и на обслуживании.

Важной особенностью систем с неограниченной очередью является то, что математическому анализу поддаются только те системы, у которых входящий поток меньше пропускной способности системы, т.е. если $\Lambda < v$.

В противном случае, т.е. если $\Lambda \geq v$, очередь будет расти до бесконечности, её состояние (длина очереди) будет зависеть от момента начала работы системы (например, с утра или с начала года) и анализ такой системы не представляет интереса.

Поскольку в системах с неограниченной очередью при $\Lambda < v$ потери вызовов отсутствуют, то обслуженная нагрузка будет равна поступающей, т.е. $Y_o = Y = \Lambda$.

9.2. Характеристики качества обслуживания для системы $M/M/v/\infty$.

а) Основной характеристикой, наиболее часто используемой для практических расчётов, является вероятность ожидания для поступившего вызова $P(\gamma > 0)$, где γ – время ожидания. По существу, это вероятность того, что вызов до начала обслуживания будет ожидать освобождения канала. Для про-

стейшего потока вызовов эта вероятность совпадает с вероятностью занятости всех каналов или с вероятностью потерь по времени:

$$P(\gamma > 0) = P_{\geq v} = P_t = \sum_{i=v}^{\infty} P_i = \frac{E_v(\Lambda)}{1 - \left(\frac{\Lambda}{v}\right)[1 - E_v(\Lambda)]} = D_v(\Lambda). \quad (9.1)$$

В данном случае суммирование проводится по всем состояниям, при которых заняты все каналы и имеется очередь любой длины: от нуля до бесконечности. Расчётное соотношение для вероятности ожидания получено Эрлангом, называется второй формулой Эрланга, обозначается как $D_v(\Lambda)$ и табулировано для практического применения. Таблицы, как и в случае первой формулы — $E_v(\Lambda)$, позволяют по любым двум из трех параметров — Λ , v , P_t — определить третий. Наиболее компактно вторая формула Эрланга описывается через первую формулу, как показано в (9.1).

Из первой и второй формул Эрланга следует, что $D_v(\Lambda) > E_v(\Lambda)$. Это означает, что при одинаковых значениях интенсивности нагрузки и числе линий вероятность ожидания в системе с ожиданием выше, чем вероятность потери вызова в системе с явными потерями.

Это объясняется тем, что в системе с потерями в момент занятости всех линий вызов теряется и никакого воздействия в дальнейшем на систему не оказывает. В системе с ожиданием при тех же условиях вызов становится в очередь и будет влиять на дальнейшее функционирование системы. При освобождении линии в системе с потерями она предоставляется поступающему вызову, а в системе с ожиданием — ожидающему. Таким образом, вновь поступившему вызову приходится становиться в очередь. Так как вызов не теряется, а лишь задерживается, то вероятность ожидания (9.1) называют вероятностью условных потерь.

Естественно возникает вопрос: какие особенности систем с ожиданием обуславливают их практическое применение и в какой области коммутационной техники такие системы целесообразно использовать? В связи с этим необходимо, прежде всего, отметить различный качественный характер явных и условных потерь. В системах с потерями часть вызовов теряется (не обслуживается), в системах с ожиданием обслуживаются все поступающие вызовы, при этом часть из них с некоторой задержкой. При малых величинах потерь, исчисляемых промилями и даже несколькими процентами, абонент не ощущает неудобств. Поэтому в области малых потерь предпочтительнее система с потерями, обладающая более высокой пропускной способностью. При больших значениях потерь система с потерями не обеспечивает должного качества обслуживания абонентов и непригодна для применения.

В случае больших потерь нельзя однозначно дать оценку системам с ожиданием. Все зависит от времени задержки обслуживания вызовов, находящихся на ожидании. Необходимо, чтобы время ожидания начала обслуживания для подавляющего большинства вызовов не вызывало неудобств у абонентов. Иными словами, системы с ожиданием следует применять в тех

случаях, когда для повышения использования коммутационных устройств целесообразно допускать большие условные потери (исчисляемые десятками процентов), если возможно при этом обеспечить мало ощутимое для абонентов время задержки начала обслуживания.

В современных телекоммуникациях, как правило, системы с потерями применяются в сетях с коммутацией каналов (отсутствует неприятное стояние абонента в очереди), а системы с ожиданием в сетях с коммутацией пакетов. Здесь абоненты не испытывают неудобств, т.к. в очереди стоят не они, а пакеты в узлах коммутации (маршрутизаторы, коммутаторы).

б) Вероятность превышения длиной очереди заданной величины n определяется с использованием второй формулы Эрланга:

$$P(S > n) = \sum_{i=v+n+1}^{\infty} P_i = (\Lambda / v)^{n+1} D_v(\Lambda).$$

в) Средняя длина очереди определяется как:

$$\bar{S} = \sum_{i=v}^{\infty} (i - v) P_i = \Lambda D_v(\Lambda) / (v - \Lambda)$$

г) Средняя длительность ожидания рассматривается в двух вариантах:

- $\bar{\gamma}_3$ - средняя длительность ожидания для задержанных вызовов, т.е. подсчитанная только среди тех, которые поступили на обслуживание после некоторого ожидания;

- $\bar{\gamma}$ - средняя длительность ожидания для всех вызовов, т.е. для любого поступившего независимо от процедуры ожидания.

$$\bar{\gamma}_3 = 1/(v - \Lambda), \quad \bar{\gamma} = \bar{\gamma}_3 D_v(\Lambda) = D_v(\Lambda) / (v - \Lambda)$$

Здесь величины $\bar{\gamma}_3$ и $\bar{\gamma}$ выражены в условных единицах времени (у.е.в.).

д) Вероятность ожидания свыше допустимого времени – это так называемые условные потери. Для упорядоченной очереди эта величина получается достаточно просто:

$$P(\gamma > t_d) = \sum_{i=v}^{\infty} P_i (\gamma > t_d) P_i = D_v(\Lambda) e^{-(v-\Lambda)t_d}$$

Здесь величина t_d тоже выражена в условных единицах времени (у.е.в.).

9.3. Расчет систем с ожиданием при ограниченном числе мест ожидания

Рассмотрим теперь случай, когда в предыдущую задачу вводятся следующие ограничения:

- число мест для ожидания конечно ($r = const$);
- вызов, поступивший, когда v линий заняты и r мест ожидания также заняты, теряется и не оказывает в дальнейшем на систему никакого влияния.

Очевидно, что такая система является комбинированной – с ожиданием и потерями. При этом условие $\Lambda < v$ теряет смысл и обслуженная нагрузка мо-

жет быть меньше поступающей, т.е. опять, как и в чистых системах с потерями, $\Lambda = Y = Y_o + Y_n$.

Для данного случая приведем только конечные соотношения.

а) Вероятность того, что поступивший в произвольный момент вызов застанет все линии занятыми, но хотя бы одно место ожидания будет свободно, т.е. наступит ожидание начала обслуживания в течение времени $\gamma > 0$, определяется формулой:

$$p(\gamma > 0) = \frac{v \left[1 - \left(\frac{\Lambda}{v} \right)^r \right]}{(v - \Lambda) / E_v(\Lambda) + \Lambda \left[1 - \left(\frac{\Lambda}{v} \right)^r \right]}.$$

б) Вероятность того, что поступивший вызов застанет все линии и места ожидания занятыми, т.е. наступит потеря вызова, определяется формулой:

$$p = \frac{(v - \Lambda) \left(\frac{\Lambda}{v} \right)^r}{\frac{v - \Lambda}{E_v(\Lambda)} + \Lambda \left(1 - \left(\frac{\Lambda}{v} \right)^r \right)}.$$

в) Средняя длительность ожидания задержанных вызовов (в у.е.в.):

$$\bar{\gamma}_z = \frac{1}{v - \Lambda} - \frac{r \left(\frac{\Lambda}{v} \right)^r}{v - v \left(\frac{\Lambda}{v} \right)^r}.$$

9.4. Расчет систем с ожиданием при ограниченном времени ожидания

Это, так называемая, модель с нетерпеливыми клиентами. Отличие этой модели заключается в следующем:

- продолжительность ожидания ограничена временем T (в у.е.в.);
- каждый ожидающий вызов по прошествии времени T покидает систему (теряется) и дальнейшего влияния на систему не оказывает.

Очевидно, что такая система также является комбинированной.

а) Вероятность того, что поступивший вызов застанет все линии занятыми, но хотя бы одно место ожидания окажется свободным, то есть наступит ожидание, определяется формулой:

$$p(\gamma > 0) = \frac{v - \Lambda e^{-(v-\Lambda)T}}{v - \Lambda / E_v(\Lambda) + \Lambda (1 - e^{-(v-\Lambda)T})}$$

б) Доля вызовов, получивших отказ в обслуживании по истечении времени ожидания T (вероятность потерь), определяется формулой:

$$P = \frac{(v - \Lambda)e^{-(v-\Lambda)T}}{(v - \Lambda)E_v(\Lambda) + \Lambda(1 - e^{-(v-\Lambda)T})}$$

в) Средняя длительность ожидания среди задержанных вызовов:

$$\bar{\gamma}_3 = \frac{1}{v - \Lambda} - \frac{Te^{-(v-\Lambda)T}}{1 - e^{-(v-\Lambda)T}}$$

10. Системы с повторными вызовами

Рассмотрим систему из v каналов, с простейшим входным потоком первичных вызовов Λ . Длительность обслуживания распределена по экспоненциальному закону со средней длительностью $h = 1$ у.е.в. Параметр потока повторных вызовов $j\beta$: где j – число источников повторных вызовов, а $\beta = h / t_{\Pi}$ – интенсивность повторения вызовов одним источником (выз/у.е.в.).

В реальной ситуации в системах с условными потерями длительность ожидания не может быть бесконечной. В системах с очередями она может ограничиваться нетерпеливыми клиентами, покидающими очередь, не дождавшись начала обслуживания, а в системах с повторениями абоненты могут по какому-либо правилу не доводить своё поведение до абсурда продолжением своих попыток до бесконечности.

Рассмотрим комбинированную систему с явными и условными потерями, работающую, например, по следующему алгоритму:

- после 1-го отказа абонент повторяет свой вызов с вероятностью H_1 , а с вероятностью $(1 - H_1)$ отказывается от повтора (в последнем случае происходит явная потеря сообщения);

- при отказе повторному вызову с вероятностью H_2 абонент повторяет свой вызов ещё раз, а с вероятностью $(1 - H_2)$ покидает систему.

Вероятности H_1 и H_2 принято называть мерой настойчивости абонентов. Функционирование такой системы зависит от конкретных значений этих вероятностей. С уменьшением H_1 и H_2 система с повторением приближается к системе с потерями и при $H_1 = 0$ становится ею. При абсолютно настойчивых абонентах ($H_1 = H_2 = 1$) параметр j (число источников повторных вызовов) может расти до бесконечности, если $\Lambda \geq v$. В системе с бесконечной очередью в этом случае до бесконечности увеличивалась длина очереди. Поэтому здесь для анализа также интересны только системы с $\Lambda < v$.

В общем случае анализ систем с повторными вызовами очень сложен, поэтому ниже приведены результаты только для одноканальной системы с абсолютно настойчивыми абонентами:

- вероятность потери первичного вызова

$$P_{B1} = \Lambda.$$

Это возможно, так как в одноканальной системе $\Lambda < 1$;

- вероятность потери повторного вызова

$$P_{в2} = (\Lambda + \beta) / (1 + \beta);$$

- вероятность потери любого вызова

$$P_v = \Lambda(1 + \beta) / (1 + \Lambda\beta);$$

- среднее число повторных попыток на одно установленное соединение Q_1 (в данном случае и на 1 первичный вызов Q_2)

$$Q_1 = Q_2 = \Lambda(1 + \beta) / (1 - \Lambda);$$

- среднее число источников, повторяющих вызовы

$$\bar{j} = \Lambda^2(1 + \beta) / \beta(1 - \Lambda).$$

11. Управляемые элементы сети связи и методы определения характеристик сети

11.1. Общие сведения

Быстрые темпы развития сетей связи и ЭВМ, большой объем оборудования, используемого на таких сетях, делают весьма важной проблему повышения эффективности функционирования сетей. Информационная сеть или ее отдельные части обслуживают потоки сообщений (телефонных, телеграфных, передачу данных и другие), характерной особенностью которых является их непрерывное изменение во времени как по величине, так и по направлениям.

Структура сети и ее технические средства (пучки каналов, коммутационное оборудование) в процессе функционирования сети развиваются и увеличиваются в объемах только через определенные промежутки времени. Таким образом, правильно спроектированная сеть, наилучшим образом обслуживающая заданные потоки сообщений, является оптимальной лишь в течение некоторого, сравнительно небольшого периода времени.

Постоянно изменяющиеся потоки сообщений приводят к тому, что возникает несоответствие между схемой потоков и структурой сети. Это несоответствие обусловлено как общим увеличением потоков, так и их перераспределением по направлениям. Общее увеличение объемов требований на передачу сообщений в течение некоторого времени компенсируется запасами технических средств, а в дальнейшем должно быть учтено очередным развитием сети.

Перераспределение потоков сообщений в пределах примерно одинаковых общих объемов приводит к уменьшению соответствия структуры сети измененному распределению потоков сообщений, снижению эффективности функционирования сети и ухудшению качества обслуживания. Таким обра-

зом, возникает проблема восстановления соответствия между распределением потоков и структурой сети, которая может быть решена за счет введения управления на сети.

Управление на сети связи может быть двух типов:

- за счет управления ресурсами сети (техническими средствами, заложенными в линиях связи и узлах коммутации);
- за счет управления потоками сообщений (изменение путей передачи сообщений практически без ограничения объема потоков).

Возможно одновременное управление и потоками, и ресурсами. Кроме того, существуют и используются такие методы управления, которые можно считать и методами управления потоками, и методами управления техническими средствами сети.

Проблема повышения эффективности сетей важна не только потому, что позволяет получить существенный экономический выигрыш. В некоторых случаях это – единственная возможность обеспечить передачу информации в требуемых объемах, т. е. без повышения эффективности сеть не может быть построена вообще. Поэтому методы анализа, синтеза и оптимизации сети приобретают исключительное значение, а в связи с этим и методы теории телетрафика, позволяющие решать отдельные задачи исследования сетей.

Необходимость управления сетью связи (изменение структуры, использование адаптивной структуры, изменение направления передачи сообщений и другие) ставит целый ряд сложных, специфических задач теории телетрафика. Большинство этих задач до сих пор еще не решено, а часть из них даже не сформулирована должным образом. Однако решение отдельных задач, связанных с управлением на сети связи, уже сейчас может быть проиллюстрировано методами теории телетрафика, и в некоторых случаях существует возможность оценить эффект от использования того или другого способа управления на сети связи.

К методам теории телетрафика, которые используются для определения отдельных характеристик сети связи, можно отнести:

- метод определения числа соединительных устройств в пучках, обслуживающих избыточную нагрузку (метод эквивалентных замен);
- методы определения характеристик управляемых процессов, возникающих как при управлении потоками, так и при управлении структурами;
- методы, используемые при определении характеристик информации, передаваемой по сети с целью управления отдельными элементами сети или сетью в целом (метод укрупнения состояний), и др.

В данном разделе рассматриваются вопросы, связанные с установлением соединений по обходным путям, которые можно считать методом управления как потоками, так и каналами на сети; охарактеризованы задачи, возникающие при динамическом управлении на сети (управлении потоками сообщений); указаны основные идеи применения кроссовой коммутации на сети связи (управление каналами связи и каналами коммутации), а также рассмот-

рен метод укрупнения состояний пучка, позволяющий определять важные характеристики управляющей информации на сети.

11.2. Расчет числа линий в обходных пучках

Установление соединений между абонентами различных АТС районированной телефонной сети осуществляется с помощью межстанционных соединительных линий (СЛ). При этом для улучшения использования соединительных линий и повышения вероятности установления соединения системы автоматической коммутации позволяют, помимо основного пути установления соединения (пути первого выбора), использовать один или несколько обходных путей (пути второго и последующего выборов). Например, на сети, содержащей четыре районных АТС, упрощенная схема которой приведена на рис. 11.1, установление соединений между абонентами, включенными в АТС *A* и АТС *B*, может производиться с использованием одной из СЛ пучка *AB* (путь первого выбора), но если все СЛ этого пучка заняты, то можно использовать обходный путь *ACB* (путь второго выбора) с занятием одной из СЛ пучка *AC* и одной из СЛ пучка *CB* (с занятием двух соединительных линий, т.е. по одной линии в каждом из пучков, составляющих обходный путь). Если же все СЛ хотя бы в одном из рассматриваемых пучков (*AC* или *CB*) заняты, то можно использовать обходный путь третьего выбора, например *ADB*.

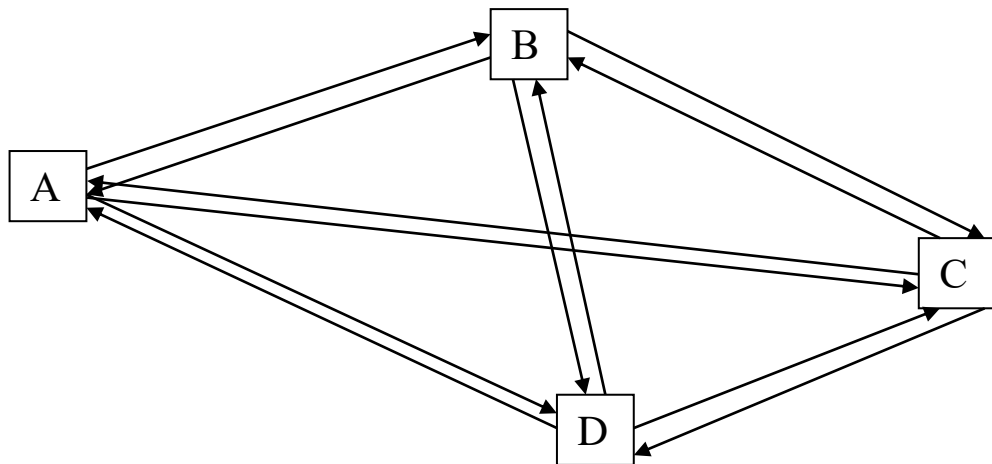


Рис. 11.1. Схема сети с четырьмя РАТС

Таким образом, основная часть телефонной нагрузки, поступающей от абонентов АТС *A* к абонентам АТС *B* (интенсивность поступающей нагрузки Y_{AB}), будет обслужена СЛ пучка *AB*, однако некоторая часть этой нагрузки в моменты занятости всех линий пучка *AB* будет предлагаться пучкам *AC* и *CB*, составляющим путь второго выбора. Эту часть нагрузки называют избыточной нагрузкой R_{AB} . Следовательно, пучок *AC* должен обслуживать как поступающую нагрузку Y_{AC} , так и избыточную нагрузку R_{AB} .

Кроме того, если СЛ пучка AC используются также и для установления соединений между АТС A и D (Y_{AD}) по обходному пути ACD в случае, когда все СЛ пучка AD (путь первого выбора) заняты, тогда пучок AC будет обслуживать поступающую нагрузку Y_{AC} , избыточную нагрузку R_{AB} (оставшуюся часть нагрузки Y_{AB} , не обслуженную пучком AB), и избыточную нагрузку R_{AD} (оставшуюся от нагрузки Y_{AD} , предложенной пучку AD).

Рассмотрим полнодоступный пучок из v линий, на первую линию которого поступает поток с интенсивностью Y . Вызовы, поступающие в моменты занятости первой линии, предлагаются для обслуживания второй и последующим линиям пучка и образуют избыточный поток для первой линии пучка. Аналогичным образом можно рассматривать избыточный поток для первых двух линий пучка, поступающий на все остальные линии, и избыточный поток для любого числа первых v_1 линий рассматриваемого пучка, поступающий на остальные $v - v_1$ линий пучка.

На рис. 11.2 приведен полнодоступный пучок, содержащий v линий, на который поступает поток $\Pi_{\text{п}}$, характеризующийся интенсивностью нагрузки Y . Избыточный поток $\Pi_{\text{и}}$ создает интенсивность нагрузки R . Если считать, что поступающий поток $\Pi_{\text{п}}$ простейший, то избыточный поток $\Pi_{\text{и}}$ не будет простейшим. Вызовы этого потока могут появиться не в любой момент рассматриваемого периода, а только в те моменты, когда все v линий пучка заняты, т. е. вызовы избыточного потока сосредоточены только на части рассматриваемого интервала времени, значит избыточный поток более концентрирован. При одной и той же нагрузке избыточный поток требует больше линий для своего обслуживания, чем простейший поток.

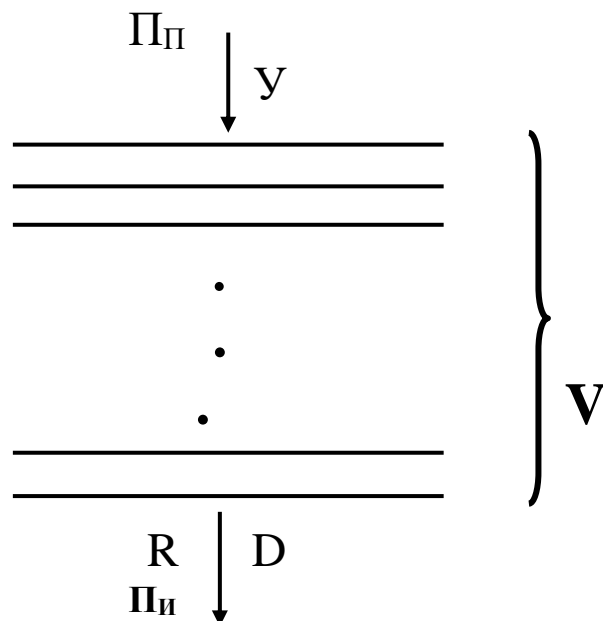


Рис. 11.2. Схема полнодоступного пучка

Для характеристики статистических (случайных) колебаний избыточного потока кроме интенсивности нагрузки, т. е. средней величины (первого

момента случайной величины), используют также дисперсию D , которая для простейшего потока равна матожиданию, а для избыточного превышает его, т.е. для избыточного потока:

$$D > R \quad (11.1)$$

11.3. Метод эквивалентных замен

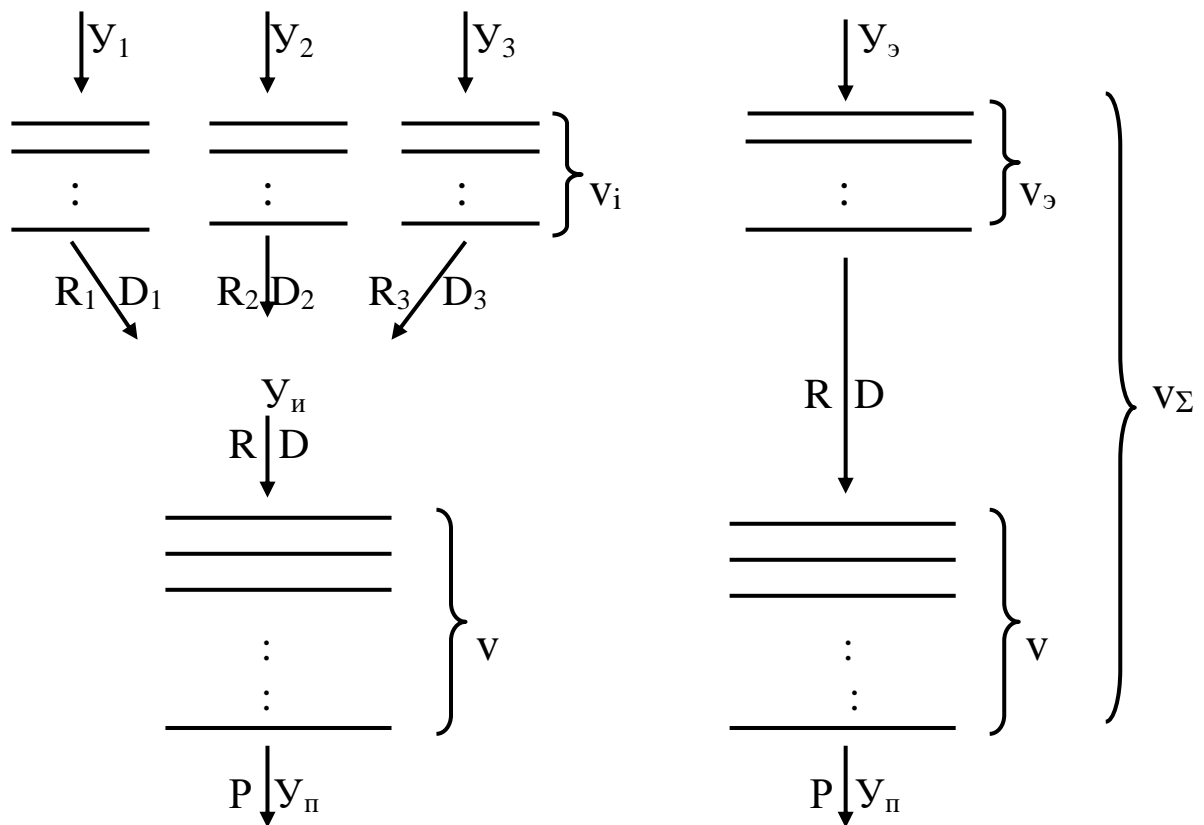
Расчёт числа линий в обходном пучке каналов рассмотрим с помощью метода Вилкинсона на примере схемы, представленной на рис. 11.3а. На пучок каналов v поступают три избыточных потока R_1, R_2 и R_3 от входящих потоков Y_1, Y_2 и Y_3 . Математические ожидания избыточных потоков определяются как функции числа каналов и величины входящего потока:

$$R_i = Y_i P_i = Y_i E_{vi}(Y_i),$$

где $E_{vi}(Y_i)$ – вероятность отказа в предоставлении канала для системы $M/M/v$.

Если дисперсия входящего простейшего потока равна его матожиданию, т.е. $D_i = R_i$, то избыточный поток обладает большей дисперсией, которая определяется как:

$$D_i = R_i \left(1 - R_i + \frac{Y_i}{V_i + 1 + R_i - Y_i} \right).$$



а) б)
Рис. 11.3. Метод эквивалентных замен

Среднее значение нагрузки и дисперсия объединенного избыточного потока равны сумме соответствующих параметров этих потоков и в общем случае (для k потоков) равны:

$$R = R_1 + R_2 + \dots + R_k, \quad (11.2)$$

$$D = D_1 + D_2 + \dots + D_k. \quad (11.3)$$

Расчет обходного пучка по рис. 11.3а заключается в определении числа линий v , если заданы характеристики потока R и D , и указана допустимая вероятность потерь $P_{\text{обх}}$ на обходном пучке.

Основная идея метода эквивалентных замен заключается в том, что схема (рис. 11.3а) заменяется на схему (рис. 11.3б) с таким расчётом, чтобы сохранить характеристики избыточного потока $Y_{\text{и}}$. При этом определяются величина некоторой эквивалентной (фиктивной) поступающей нагрузки $Y_{\text{э}}$, простейшего потока вызовов и некоторое фиктивное число линий $v_{\text{э}}$ в предварительном пучке такие, что на выходе пучка $v_{\text{э}}$ мы можем получить такие же значения для R и D , какие были получены нами по ф-лам (11.2, 11.3) для схемы рис. 11.3а.

С другой стороны, предварительный пучок $v_{\text{э}}$ и рассчитываемый пучок v составляют один общий полнодоступный пучок линий v_{Σ} , для которого найдена интенсивность поступающей нагрузки $Y_{\text{э}}$ и можно определить вероятность потерь $P_{\text{экр}}$.

Для определения значений $Y_{\text{э}}$ и $v_{\text{э}}$ можно воспользоваться приближёнными формулами:

$$Y_{\text{э}} = D + 3 \frac{D}{R} \left(\frac{D}{R} - 1 \right) \quad \text{и} \quad v_{\text{э}} = Y_{\text{э}} \frac{R^2 + D}{R^2 + D - R} - R - 1.$$

Величина потерь $P_{\text{экр}}$ определяется по очевидному соотношению как:

$$P_{\text{экр}} = R P_{\text{обх}} / Y_{\text{э}}.$$

Далее, пользуясь первой формулой Эрланга, можно определить в схеме рис. 11.3б суммарное число линий в полнодоступном пучке, обслуживающем нагрузку $Y_{\text{э}}$ с потерями $P_{\text{экр}}$, т. е. определить сумму

$$v_{\Sigma} = v_{\text{э}} + v.$$

И наконец, искомое число линий, в расчетном пучке, одинаковое для схем рис. 11.3а и 11.3б, определится как

$$v = v_{\Sigma} - v_{\text{э}}$$

Числа линий v_{Σ} и $v_{\text{э}}$ могут быть дробными, и их следует использовать в таком виде до получения результатов для v . Значение v целесообразно округлять до целого числа в сторону увеличения.

Отметим, что среди потоков в схеме рис. 11.3а, поступающих на обходной пучок, не обязательно, чтобы все были избыточными. Метод расчёта не

изменится, если один или более поступающих потоков будут простейшими. Например, если необходимо учесть собственную нагрузку обходного пучка..

12. Анализ буферной памяти узла коммутации

12.1. Процесс буферизации в узле коммутации и схемы организации буферной памяти

Современные узлы коммутации сообщений (пакетов) используют для приема, передачи и обработки сообщений ограниченную буферную память. Структура памяти, в которой скапливаются сообщения от многочисленных одновременно работающих каналов связи, в значительной мере определяет пропускную способность узла.

Функционирование УК включает многоэтапный процесс буферизации сообщений (пакетов). Несмотря на то, что господствующим методом коммутации в современных сетях является метод коммутации пакетов и что коммутация пакетов в перспективе будет единственным способом коммутации в сетях NGN (движение к All-IP), ещё существуют сети, например, телеграфные, в которых применяется метод коммутации сообщений. В данном разделе процессы буферизации в узлах коммутации рассматриваются на примере коммутации сообщений различной длины. Однако, представленная ниже методика анализа в равной степени подходит и для коммутации пакетов.

Входящее в узел сообщение записывается в буферную память со скоростью канала передачи данных. Затем обрабатывается процессором узла, выполняющим требуемое обслуживание сообщения и определяющим направление выдачи, и передается по каналу в выходящем из УК направлении. Копия принятого в УК сообщения хранится в буферной памяти до получения квитанции (логического подтверждения) о безошибочной доставке сообщения адресату или промежуточному узлу. Отсутствие квитанции в течение определенного интервала времени (time-out) вынуждает повторить передачу сообщения по выходящему каналу в направлении адресата. Это обеспечивает логическую защиту от ошибок в передаче и возможной потери сообщения из-за отсутствия свободных буферов в принимающем УК. При отсутствии места в буферной памяти входящее в УК сообщение получает отказ. Тогда сообщение направляется по другому маршруту или запрос периодически повторяется.

Этапы передачи по каналам связи УК входящих и выходящих сообщений характеризуются особенностями линейных протоколов, определяющих, по существу, эффективную скорость передачи по каналам. Особенности этапов квитирования и time-out определяются сетевыми протоколами межузловых связей и сквозного взаимодействия источник-адресат. Методы доставки пакетов (датаграммы, виртуальные соединения и т. д.) существенно влияют на время хранения копии сообщения (пакета) в буферной памяти УК-отправителя.

Таким образом, поступившее в УК сообщение занимает буферную память в течение следующих интервалов времени:

- запись со скоростью канала связи;
- ожидание в очереди на обработку в процессоре;
- обработка в процессоре;
- ожидание освобождения канала в направлении выхода из УК;
- передача из УК по выходному каналу связи;
- ожидание квитанции об удачной доставке с возможным повторением по истечении интервала time-out.

Сообщение (пакет) освобождает занимаемую им память после выполнения всей последовательности операций. Конкретный набор этапов зависит от реализации УК и его функций, а также от используемых сетевых технологий. Описанный многоэтапный процесс адекватно отражает известные процедуры коммутации сообщений (пакетов) для различных структур буферной памяти узла.

В узле коммутации с большим числом каналов (направлений) связи буферная память разбивается на участки (буферы) для параллельного обслуживания нескольких сообщений (пакетов). Способы разбиения разнообразны и зависят от метода организации памяти. Различают статические и динамические методы.

В первом случае за каждым каналом жестко закрепляется один или несколько буферов, в каждый из которых можно записать сообщение максимальной длины. Этот метод отличает неэффективное использование буферной памяти. При более рациональных динамических методах буферная память является общим для всех каналов ресурсом. Поиск свободного буфера при этом осуществляется лишь в момент появления запроса на него. Каждый буфер может быть выделен для любого канала. Размер буфера может определяться сообщением максимальной длины или быть меньше его. В последнем случае для приема сообщения выделяется цепочка буферов и память при большом разбросе длин сообщений используется эффективнее.

Опишем более подробно схему организации памяти с буферами на максимальную длину сообщений. В простейшем случае такая память организуется из идентичных равнодоступных буферов. Эта схема широко применяется в узлах коммутации пакетов и сообщений. При этом многоэтапный процесс буферизации может быть формализован в виде однородной сети МО. Выделение на входе узла классов сообщений с заранее известной длиной позволяет организовать память секциями буферов (секция на класс). Объем идентичных буферов секции определяется максимальной длиной сообщений соответствующего класса. Такая структура памяти, по сравнению с простейшей, позволяет экономить необходимую буферную память.

Аппаратура передачи данных зачастую не формирует сведений о длине передаваемых сообщений. В этом случае экономия памяти секционной структуры может достигаться за счет переписи принятых в общедоступный пул сообщений в секции меньших размеров. Объем памяти секционной

структуры может быть определен с помощью моделей сетей МО с несколькими классами сообщений.

Рассмотрим, наконец, схему организации динамической памяти с цепочкой буферов. Различают два вида такой памяти: цепочки буферов, резервируемые в момент поступления сообщения известной длины, и цепочки, формируемые по мере заполнения сообщением очередного буфера.

В первом случае сведения о длине передаются в начале сообщения. В момент прихода в узел оно может получить отказ в приеме при отсутствии соответствующего количества буферов. Поведение УК при этом описывается сетью МО с несколькими классами сообщений, где класс сообщения определяется количеством требуемых буферов.

Схема организации памяти второго типа не требует сведений о длине сообщений. Память при этом используется эффективно, однако возможны отказы в приеме из-за отсутствия свободных дополнительных буферов в процессе приема сообщения. Функционирование УК в этом случае описывается сетью МО с классами сообщений, также определяемыми количеством требуемых сообщений буферов, но с более сложным взаимодействием между классами.

Отметим, что анализ описанных схем организации буферной памяти УК с учетом многоэтапного процесса буферизации приводит к необходимости расчета сетей МО с одним или несколькими классами сообщений, что определяет обобщенный подход к исследованию вопросов буферизации в узлах компьютерных сетей.

Ниже в постановочном плане рассмотрены алгоритмы решения задач для двух различных схем организации буферной памяти УК:

- однородный пул равнодоступных буферов;
- пул с разделением на классы по длине сообщений.

12.2. Анализ однородного пула равнодоступных буферов

Рассмотрим модель УК с простейшей схемой организации памяти, состоящей из буферов на максимальную длину сообщений. Формализация многоэтапного процесса буферизации для этой схемы наиболее проста и в то же время включает общие для всех схем этапы.

Для расчета буферной памяти УК может быть формально представлен в виде открытой сети МО с общим накопителем в N буферов, отказами, блокировкой и пуассоновским потоком поступлений сообщений в накопитель с интенсивностью Λ (рис. 12.1).

Первым обслуживающим блоком в сети является многолинейный центр обслуживания «Память». Число обслуживающих приборов здесь равно числу мест накопителя N ; очередь отсутствует. Длительность обслуживания приборов в этом центре равна времени передачи входящего сообщения по каналу связи.

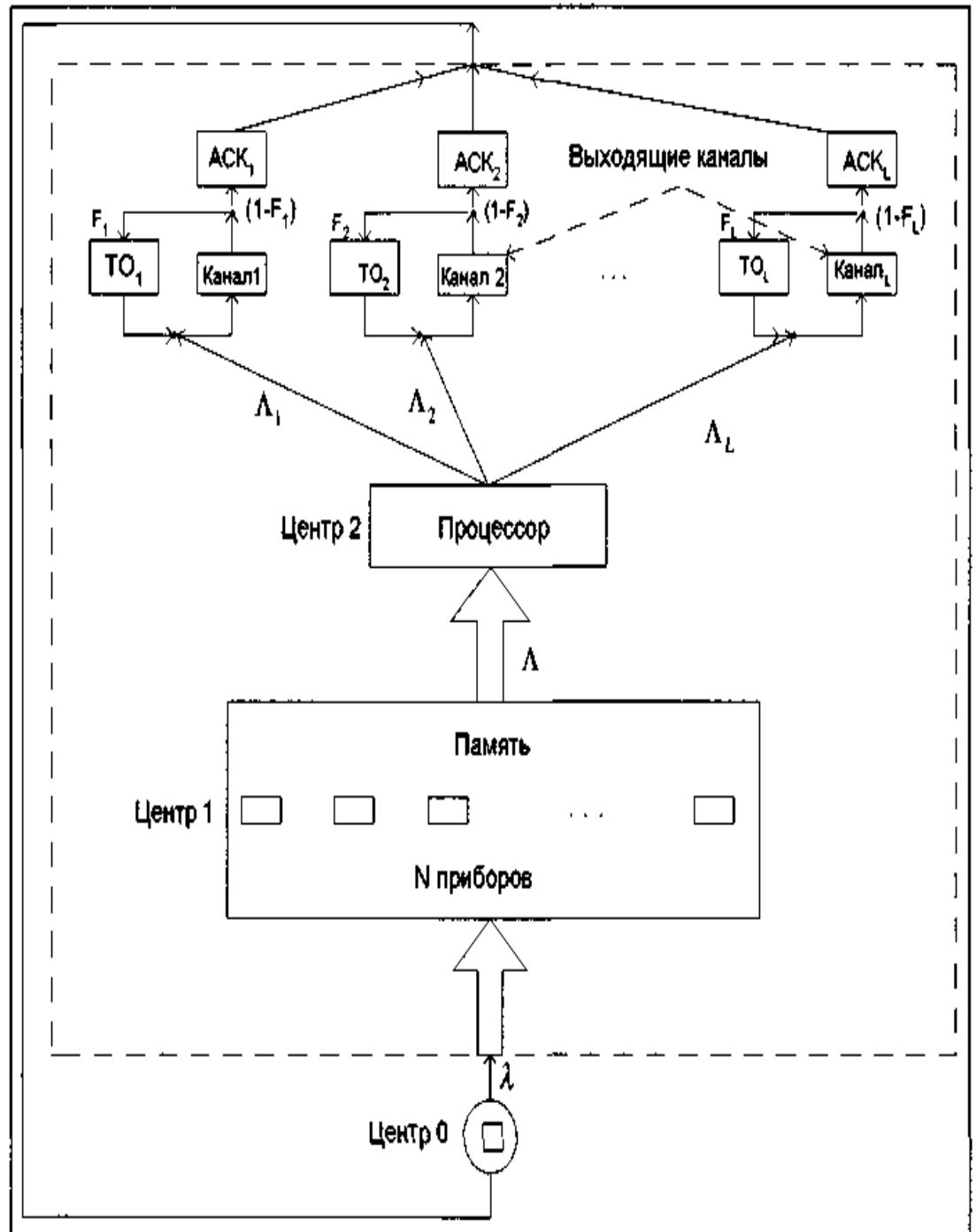


Рис. 12.1. Схема буферной памяти в виде открытой сети МО

Во втором однолинейном центре «Процессор» длительность обслуживания равна времени обработки сообщения в процессоре УК. Обслуживание сообщений в этом центре осуществляется в соответствии с дисциплиной FIFO (первым пришёл – первым обслужен).

Выделим группу из L центров, формализующих работу L выходящих каналов УК. Каждое сообщение, обслуженное во втором центре, с вероятно-

стью P_j , $j = \overline{1, L}$ поступает на обработку в j -й центр группы выходящих каналов ($P_1 + P_2 + \dots + P_L = 1$). Дисциплина обслуживания очередей в центрах этой группы - FIFO. Сообщения, прошедшие обслуживание в j -м центре упомянутой группы, с вероятностью $1 - F_j$ поступают в соответствующий центр АСК, нормализующий ожидание логического подтверждения передачи по каналу j . В противном случае с вероятностью F_j эти сообщения поступают в центр TO_j , моделирующий задержку time-out, после которой осуществляется повторное обслуживание сообщения в j -м центре группы выходящих каналов. Вероятность F_j , $j = \overline{1, L}$ определяет условия неудачной передачи сообщения соседнему узлу или адресату, включая отсутствие в последних свободной буферной памяти.

Прибор центра «Память» после окончания обслуживания сообщения блокируется (не освобождается) до тех пор, пока не завершится обслуживание сообщения в одном из центров АСК. Сообщения, поступающие в рассматриваемую сеть МО, могут получать отказ, если в момент их появления все приборы центра «Память» блокированы или заняты. Интенсивность входящего в сеть потока зависит от числа занятых мест накопителя n и имеет вид:

$$\Lambda(n) = \begin{cases} 0, & \text{при } n = N \\ \Lambda, & \text{при } n < N \end{cases}$$

Математическая модель УК позволяет определить зависимость вероятности отказа от количества мест (буферов) накопителя. По этой зависимости, исходя из допустимой вероятности отказов входящим в узел сообщениям, легко определить объем буферной памяти.

Анализ описанной МО в общем случае при произвольных распределениях длительности обслуживания сообщений в центрах с дисциплиной FIFO можно осуществить с помощью приближенных методов.

12.3. Анализ пула с разделением на классы по длине сообщений.

Как отмечалось ранее, большой разброс длин сообщений делает нерациональным использование однородного пула равнодоступных буферов. Создаваемые при этом буфера на максимальную длину сообщения приводят к чрезмерно большой буферной памяти УК. Вместе с тем разделение входящего в УК потока сообщений на классы по длине сообщений позволяет организовать более экономную секционную структуру буферной памяти. Каждому классу сообщений ставится в соответствие секция идентичных буферов, объем которой соответствует максимальной длине сообщения одного класса. При этом наиболее существенная экономия достигается для «полимодаль-

ных» распределений длин входящих сообщений, когда длины концентрируются около нескольких значений.

Возможность выделения указанных классов на входе УК определяется наличием информации о длине передаваемых сообщений. Различают два случая:

- длина передается в начале самого сообщения или определяется в момент поступления сообщения в УК иным образом;
- сведения о длине отсутствуют в формате сообщения и не формируются аппаратурой передачи данных.

В последнем случае прием сообщений всех классов осуществляется в секцию с буферами на максимально возможную длину сообщений. По окончании ввода сообщений оно переписывается в секцию допустимо меньших буферов соответствующего класса.

В рассматриваемом случае УК может быть формально представлен в виде открытой сети МО с R классами сообщений, отказами, блокировкой и пуассоновскими потоками поступлений сообщений соответствующих классов с интенсивностями $\Lambda_r (r = \overline{1, R})$. Основное отличие этой модели от описанной в предыдущем пункте состоит в замене одного центра «Память» на R аналогичных центров. Число приборов в каждом r -м центре равно N_r ; очередь отсутствует. Длительность обслуживания прибором в таком центре равна времени передачи входящего сообщения класса r по каналу связи. Остальные центры сохраняют структуру, определенную в предыдущем пункте.

Вероятность перехода сообщений между центрами сети зависит от класса сообщения. Сообщение класса $r (r = \overline{1, R})$, обслуженное в центре «Процессор», с вероятностью $P_j(R) (j = \overline{1, L})$ поступает в j -й центр группы входящих каналов. По окончании обслуживания этим центром оно с вероятностью $1 - F_j(r)$ поступает в j -й центр АСК.

Место, занятое сообщением класса r блокируется до завершения обслуживания сообщения в одном из центров АСК сети. Сообщение класса r при поступлении в сеть МО получает отказ, если все места накопителя r заняты. При этом зависимость интенсивности входящего потока класса r от числа занятых мест n_r соответствующего накопителя имеет вид:

$$\Lambda_r(n_r) = \begin{cases} 0, & \text{при } n_r = N_r, \\ \Lambda_r, & \text{при } n_r < N_r. \end{cases}$$

Описанная сеть МО имеет мультипликативную форму решения в предположении экспоненциального распределения времени обслуживания в узлах с дисциплиной в очереди FIFO. При этом среднее время обслуживания сообщений всех классов одинаковы.

13. Расчёт характеристик узлов коммутации в сетях

Основными характеристиками определяющими качество функционирования сетей передачи данных, как и вообще любых сетей с коммутацией пакетов, являются длительности задержек пакетов, включая джиттер, и вероятности потерь пакетов. Источниками задержек и потерь являются узлы коммутации и каналы связи, через которые эти пакеты передаются. Ниже приводятся методики расчётов основных вероятностно-временных характеристик для узлов коммутации пакетов.

13.1. Расчет длительности задержек в узле коммутации пакетов

Рассматривается задача расчета средней длительности задержек в узле коммутации пакетов. Термин «узел коммутации пакетов» означает здесь и концентратор (статистический мультиплексор), и узел виртуальной коммутации пакетов (сети X.25, Frame Relay, сети АТМ) и маршрутизатор (сети IP). Узел коммутации пакетов может быть представлен в виде элемента с множеством входных каналов и одним выходным каналом (концентратор) или элемента с множеством входных и выходных каналов (коммутатор, маршрутизатор). С использованием символики Кендалла такие сетевые элементы могут быть представлены системами массового обслуживания вида $G/G/1$ или $G/G/n$ (произвольные вероятностные распределения, описывающие и входящий поток заявок (в нашем случае - пакетов или протокольных блоков), и время их обслуживания. Отметим, что при анализе узлов коммутации пакетов часто используются модели с одним обслуживающим прибором, то есть системы $G/G/1$.

Средняя длина очереди в системе $M/G/1$ (пуассоновский поток пакетов на входе, произвольное распределение времени обслуживания) при бесконечном размере буфера рассчитывается по классической формуле Хинчина-Полячека:

$$\bar{q} = \rho + \rho^2 \frac{1+C_s^2}{2(1-\rho)}, \quad \rho < 1 \quad (13.1)$$

где $\rho = \frac{\lambda}{\mu}$ - нагрузка системы массового обслуживания (отношение интенсивности входящего потока заявок к интенсивности их обслуживания)

$C_s^2 = \frac{D(t_s)}{(\bar{t}_s)^2}$ - квадратичный коэффициент вариации распределения времени обслуживания;

$D(t_s)$ - дисперсия распределения времени обслуживания;

\bar{t}_s - среднее время обслуживания протокольного блока (датаграммы, пакета, кадра, ячейки) в системе.

Для определения средней длительности задержки в системе $M/G/1$ воспользуемся формулой Литтла:

$$\bar{q} = \lambda \bar{t}_q$$

где \bar{q} – средняя длина очереди в рассматриваемой системе.

Тогда средняя длительность задержки определится как:

$$\bar{t}_q = \frac{\bar{q}}{\lambda} = \bar{t}_s \left[1 + \rho \frac{1+C_s^2}{2(1-\rho)} \right] \quad (13.2)$$

Для расчета средней длины очереди и средней длительности задержки необходимо знать значения дисперсии и математического ожидания (или коэффициента вариации) распределения времени обслуживания пакета или любого другого протокольного блока (время обслуживания пропорционально длине протокольного блока). В табл. 13.1 приведены выражения для расчета квадратичных коэффициентов вариации некоторых распределений, применяемых при оценке средней длительности задержки в сетях Интернет.

Гиперэкспоненциальное распределение формируется путем суммирования нескольких экспоненциальных распределений; каждое входящее в сумму распределение при этом взвешивается, и сумма весов равна единице. В табл. 13.1 гиперэкспоненциальное распределение представлено для случая, когда результирующее распределение содержит две составляющие. Вес одной составляющей равен S , вес второй составляющей равен $(1-S)$.

Геометрическое распределение, представляющее в таблице единственное дискретное распределение вероятностей, используется применительно к некоторым приложениям для описания распределении времени пребывания системы в определенном состоянии. Геометрическое распределение характеризуется вероятностью p_i пребывания системы в этом состоянии и вероятностью $1 - p_i$ нахождения системы в другом состоянии.

Параметры систем вида $G/G/1$ с бесконечной памятью не могут быть рассчитаны точно при распределениях параметров входящих потоков, отличных от пуассоновского. Однако существует набор приближенных формул, позволяющих рассчитать очереди и задержки. Ниже приведены формулы для расчета средней длины очереди в системе $G/G/1$, откуда легко может быть получена средняя длительность задержки:

$$\bar{q}_1 = \rho \left[1 + \frac{\rho(C_a^2 + C_s^2)}{2(1-\rho)} \right] \quad (13.3)$$

$$\bar{q}_2 = \rho \left[\frac{1}{2} + \frac{\rho C_a^2 + C_s^2}{2(1-\rho)} \right] \quad (13.4)$$

$$\bar{q}_3 = \frac{\rho C_a^2 + C_s^2}{2(1-\rho)} \quad (13.5)$$

где C_a^2 и C_s^2 - квадратичные коэффициенты распределения входящего потока протокольных блоков и времени их обслуживания, соответственно.

Таблица 13.1. Квадратичные коэффициенты вариации для некоторых распределений

Распределение	Коэффициент C^2
Экспоненциальное	$C^2 = 1$
Эрланга	$C^2 = \frac{1}{k}$ (k -порядок распределении Эрланга)
Гиперэкспоненциальное (H)	$\frac{1-2S+2S^2}{2S(1-S)}, 0 < S < \frac{1}{2}$ (s – параметр гиперэкспоненциального распределения для суммы двух экспонент)
Геометрическое (Geom)	$C^2 = p_i, 0 < p_i < 1$ p_i – параметр геометрического распределения
Постоянное время обслуживания заявки (D)	$C^2 = 0$

Из формул для оценки средних длин очередей (задержек) видно, что в знаменателе каждой формулы присутствует множитель $(1-\rho)$, который является полюсом уравнения.

При приближении ρ к единице полюс стремится к нулю и кривые для очередей (задержек) стремятся к бесконечности (рис. 13.1). Это явление должно учитываться при выборе нагрузки, чтобы обеспечить границы длительности задержек в соответствии с нормами качества обслуживания,

Приближение (13.3) сводится к формуле Хинчина-Полячека, то есть является точным для системы $M/G/1$. Использование той или иной приближенной формулы для расчета очереди определяется тем, насколько распределение входящего потока отличается от пуассоновского, а также от нагрузки обслуживающего устройства ρ .

13.2. Расчет вероятности потерь в узле коммутации пакетов

Еще одним важным параметром QoS в сетях передачи данных является вероятность потерь пакетов. Имеется ряд причин, из-за которых пакеты не доставляются в пункт назначения.

Среди основных причин отметим искажение пакетов в процессе передачи через сеть, превышение «времени жизни» пакетов, а также отброс пакетов в узлах при отсутствии свободного места в буферном накопителе узла.

Последнее явление встречается в том случае, если накопитель имеет конечную емкость памяти. Вероятность потерь определяется как вероятность переполнения буферного накопителя.

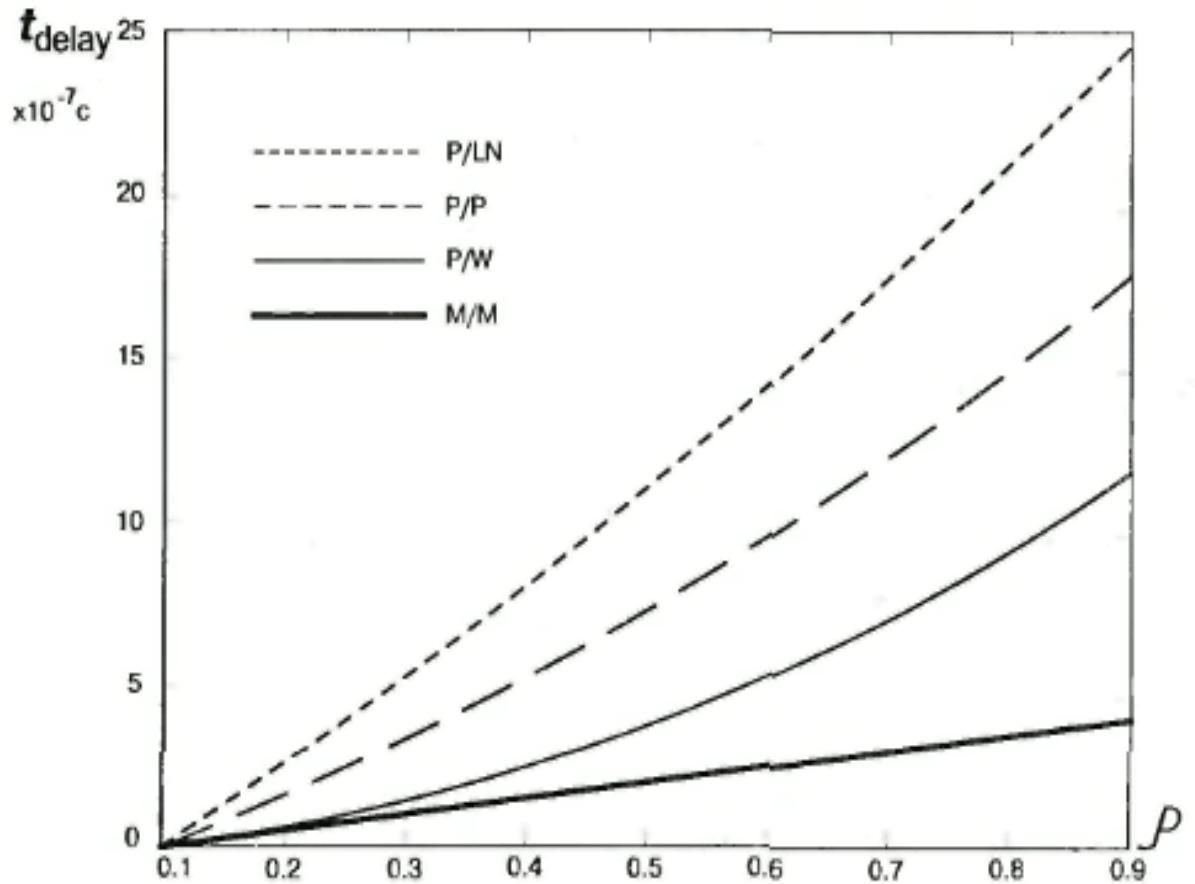


Рис. 13.1. Зависимость длительности задержки в узле коммутации от нагрузки при разных видах распределения характеристик входящего потока и времени обслуживания

В данном разделе рассматривается задача расчета вероятности переполнения памяти в узле, который в общем виде описывается системой массового обслуживания вида $G/G/1/N$. Начнем с модели простейшей системы с пуассоновским входящим потоком и экспоненциальным распределением времени обслуживания, а затем рассмотрим более общие модели системы массового обслуживания.

Система $M/M/1/N$. Вероятность переполнения памяти определяется на основе процессов гибели и размножения и равна:

$$P_{loss} = \frac{1 - \rho}{1 - \rho^{N+1}} \rho^N \quad (13.6)$$

Очевидно, что при значениях $\rho^N \ll 1$ для системы $M/M/1/N$ может быть использована следующая аппроксимация:

$$P_{loss} \approx \rho^N \quad (13.7)$$

На рис. 13.2 представлены зависимости вероятности переполнения памяти P_{loss} от размера памяти накопителя N , измеряемого в числе протокольных блоков (пакетов, кадров, ячеек), при разных значениях нагрузки ρ . Эти зависимости рассчитаны по формуле (13.7). Заметим, что для фиксированного размера буфера вероятность переполнения возрастает по мере увеличения нагрузки ρ .

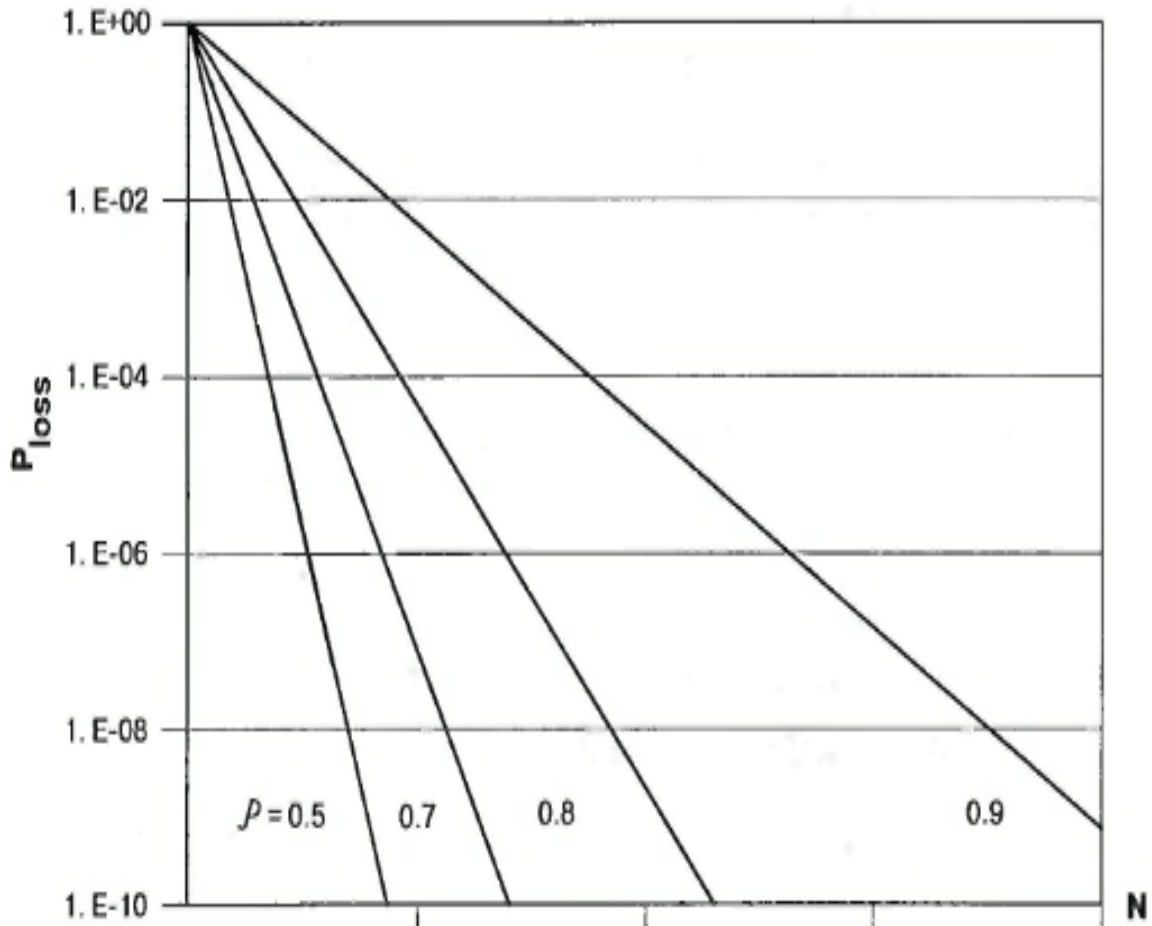


Рис. 13.2. Вероятность потерь как функция размера буферного накопителя при разных значениях ρ

При заданной вероятности переполнения памяти существует максимальное значение нагрузки узла, при которой система с очередями удовлетворяет требованиям к вероятности потерь, определяемых нормами для этой вероятности.

Из уравнения (13.7) можно также получить необходимый размер буфера в узле, исходя из вероятности потерь. Решение уравнения относительно емкости буфера N выражается следующей формулой:

$$N \approx \frac{\ln(P_{loss})}{\ln(\rho)} \quad (13.8)$$

Система G/G/1/N. Получение точных решений в замкнутой форме для систем такого типа при известных распределениях входящего потока и времени обслуживания, особенно при конечной емкости накопителя, сопряжено со значительными трудностями. Более эффективным является использование приближенных, но простых в применении оценок, базирующихся на квадратичных коэффициентах вариации входящего потока и времени обслуживания. Приближенная формула для оценки вероятности потерь в системе, если эти параметры распределений входящего потока и времени обслуживания известны, имеет следующий вид:

$$P_{loss} \approx \frac{1 - \rho}{1 - \rho \frac{c_a^2 + c_s^2}{2}^{N+1}} \rho \frac{c_a^2 + c_s^2}{2}^N \quad (13.9)$$

Зная распределения входящего потока и времени обслуживания и, таким образом, получив значения квадратичных коэффициентов вариации, можно рассчитать вероятность потерь в довольно сложной системе, допуская, что эти оценки будут приближенными.

14. Анализ мультисервисных сетей

14.1. Особенности анализа мультимедийного трафика в сетях IP

Сегодня весь сетевой трафик Интернет практически можно разделить на два класса - трафик, передаваемый под управлением протокола TCP и трафик, передаваемый под управлением протокола UDP.

Пропорции трафика TCP и UDP изменялись очень мало в течение последних 5-7 лет. Примерно 90% трафика передается через соединения TCP. Приложения, влияющие на рост трафика TCP, развиваются очень быстро, в первую очередь, благодаря разнообразным web-приложениям и одноранговым (peer-to-peer) межсетевым соединениям. В то же время наблюдается рост объемов и трафика UDP в связи с нарастающей популярностью новых приложений, таких как VoIP, IPTV и др. Примерные объемы трафика UDP сегодня составляют около 9%, однако следует ожидать существенный рост этого класса трафика в ближайшие годы.

В сетях IP присутствует также *трафик управления*, который формируется различными протоколами сигнализации и управления сетью. Хотя обработка трафика управления необходима для нормального функционирования сети, его объем относительно мал (1 – 1,5%) и не влияет существенно на соотношение объемов трафиков в сети.

Модели сетей массового обслуживания (МО) применяются для анализа характеристик протоколов практически всех уровней (в первую очередь второго, третьего и четвертого). На канальном уровне эти модели используются для определения эффективной скорости передачи данных. При анализе сквозных протоколов (транспортный уровень) модели теории сетей МО позволяют находить межконцевую задержку сообщений (пакетов), определять

параметры управления потоками и т.д. Сетевые модели отдельных компонент компьютерной сети адекватно отражают многоэтапный процесс обработки сообщений (пакетов) в этих устройствах, позволяя не только рассчитывать характеристики, но и осуществлять выбор различных параметров, например объемов буферной памяти узлов коммутации. Сложную структуру отдельных узлов необходимо учитывать и при расчете транспортной сети в целом. В то же время необходимость решения оптимизационных задач (выбор топологии и пропускных способностей каналов связи, отыскание оптимальных маршрутов и т. д.) требует применения упрощенных моделей сетей МО, позволяющих находить явный вид целевой функции, в качестве которой обычно используется время задержки сообщений (пакетов). Такой упрощенный подход требует дальнейшего уточнения характеристик сети с помощью более реальных моделей, учитывающих особенности сетевых протоколов. Это приводит к итерационной процедуре проектирования компьютерных сетей.

При анализе мультисервисных сетей в зависимости от вида транспортного протокола, в сетях Интернет принято различать трафик трех основных типов: эластичный, потоковый и реального времени.

Термин «*эластичный*» применяется к трафику, создаваемому при передаче данных под управлением протокола ТСП, и его название связано с тем, что скорость передачи может изменяться в широких пределах в ответ на изменения нагрузки сети. Трафик этого вида чувствителен к потерям и мало критичен относительно задержек.

Потоковый трафик порождается приложениями, связанными с передачей аудио и видеoinформации. Эти приложения генерируют потоки пакетов, имеющие определенную скорость передачи, которая должна быть сохранена во время сеанса связи путем ограничения задержек, но при этом допустимы более длительные, по сравнению с трафиком реального времени, задержки, и этот тип трафика относительно малочувствителен к потерям.

Трафик реального времени допускает относительно небольшие длительности задержек и малочувствителен к потерям. Трафик этого типа создается в системах IP-телефонии и видеоконференц-связи. И потоковый трафик, и трафик реального времени передаются под управлением протокола UDP.

В классических сетях IP присутствовал только эластичный трафик, который обслуживался по принципу best effort. Для современных сетей IP, где имеется трафик всех трех типов, требуется широкий набор показателей параметров качества обслуживания, начиная от параметров уровня best effort и заканчивая параметрами, соответствующими трафику реального времени.

Предположением, необходимым для возможности использования аналитических моделей сетей МО, является предположение о независимости, суть которого состоит в том, что времена передачи сообщений (пакета) по разным каналам связи предполагаются независимыми случайными величинами. В то же время очевидно, что длительности обслуживания пакетов одного сообщения в разных каналах пропорциональны длине этого сообщения

и, следовательно, зависимы. Дополнительные зависимости вносятся процессами сборки и разборки сообщений на пакеты. Тем не менее, многочисленные сравнения результатов аналитического моделирования с помощью сетей МО и результатов имитационного моделирования или измерений на реальных сетях показали, что постулирование независимости не вносит существенных погрешностей. Указанный факт объясняется следующими причинами:

- нарушением корреляции между длительностями обслуживания и длинами сообщений (пакетов) из-за восстановления после ошибок (например, на канальном уровне);
- объединением различных независимых входящих потоков в одном исходящем канале;
- использованием в основном средних характеристик, которые менее чувствительны к описанной выше зависимости.

Значительное влияние на характеристики сети пакетной коммутации оказывают способы доставки пакетов (датаграммы или виртуальные каналы) и методы маршрутизации. Моделирование различных способов доставки пакетов и методов маршрутизации осуществляется путем выбора соответствующего типа сети МО и матрицы маршрутов $\|P_{ij}\|$, структура которой учитывает и топологию сети. Например, в качестве модели датаграммной сети со случайной процедурой выбора маршрутов часто применяется открытая или замкнутая однородная сеть МО, а для моделирования постоянных виртуальных каналов (с возможным наличием альтернативных путей) и фиксированной маршрутизации - модель сети МО с несколькими классами сообщений. При этом учет служебных сообщений может быть осуществлен различными способами в зависимости от используемых протоколов: либо увеличением длительности обслуживания сообщений в модели сети МО, либо введением дополнительного приоритетного класса служебных сообщений (длина служебных сообщений значительно меньше пакета данных).

В заключение отметим, что модели сетей МО (как и любые математические модели) не в состоянии полностью отразить сложные и многообразные информационные процессы в компьютерных сетях и, кроме того, их использование обусловлено рядом предположений (таких, как предположение о независимости).

Трудности анализа сетевых структур проиллюстрируем на примере анализа пропускной способности сети в самом простом варианте – двух узловая сеть с двумя одинаковыми каналами, например, в 100 Мбит/с (Рис. 14.1). Даже формулировка этой задачи в постановочном плане – не однозначна.

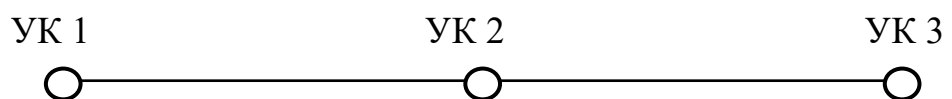


Рис. 14.1. Схема трёх узловой сети

Какой поток может пропустить эта сеть? Можно ли сопоставлять с пропускной способностью объём передаваемого этой сетью трафика. Даже не касаясь очень важных вопросов о дисциплине обслуживания (потери или очереди) и о качестве обслуживания (QoS, SLA), этот вопрос не однозначен и не может рассматриваться отдельно от схемы распределения потоков. Например, приведённая на рис.14.1 сеть для потоков, заданных матрицами C_1 или C_2 будет пропускать различные суммарные потоки.

$$C_1 \text{ [Мбит/с]} = \begin{vmatrix} - & 70 & 0 \\ 70 & - & 70 \\ 0 & 70 & - \end{vmatrix} \quad C_2 \text{ [Мбит/с]} = \begin{vmatrix} - & 0 & 70 \\ 0 & - & 0 \\ 70 & 0 & - \end{vmatrix}$$

Очевидно, что для тяготения C_1 битовый поток по сети в целом будет вдвое большим, чем для C_2 , а при стремлении степени загрузки каналов $\rho \rightarrow 1$, это превышение может стать ещё большим за счёт возможных проблем в узле УК 2 для потоков, заданных матрицей C_2 . А если исходящие из узлов потоки будут распределяться по другим узлам в какой-либо пропорции и даже могут быть асимметричными как, например, в матрице тяготений C_3 , то задача ещё более усложняется. Таким образом, даже для трёх узловой сети в зависимости от конкретной матрицы тяготений суммарная пропускная способность сети будет различной.

$$C_3 \text{ [Мбит/с]} = \begin{vmatrix} - & 50 & 20 \\ 100 & - & 40 \\ 45 & 25 & - \end{vmatrix}$$

В теории графов есть теорема о максимальном потоке: максимальный поток между двумя вершинами равен минимальному потоку среди всех сечений, разделяющих эти вершины. Эта теорема столь же очевидна сколь и бесполезна, так как справедлива только для двухполюсной сети, т.е. для сети, в которой только один узел отправитель и только один узел получатель.

Достаточно приближённые аналитические модели для многополюсных сетей ещё не созданы, поэтому большинство расчётов производятся на уровне математических ожиданий информационных потоков по заранее установленным маршрутам.

Рассмотренные в предыдущем разделе примеры вычислений длительности задержек и вероятности потерь позволяют провести оценки для изолированных сетевых узлов. Вычисление сквозной вероятности потерь в сети, учитывающей только потери в узлах из-за переполнения памяти, является более сложной задачей. Однако, приняв гипотезу Л. Клейнрока о независимости

узлов в сети Интернет, можно свести расчет потерь в сети, содержащей K узлов, к простой формуле:

$$P_{loss_{net}} = 1 - \prod_{k=1}^K 1 - P_{loss_k} \quad (14.1)$$

где $P_{loss_{net}}$ - вероятность сквозных потерь для сети,

P_{loss_k} - вероятность потерь в узле с номером k .

Получения длительностей сквозных задержек в сети рассмотрено ниже.

14.2. Расчёт времени задержки доставки пакетов в сетях

Для построения модели сети МО, описывающей функционирование базовой сети, которая включает M каналов передачи данных и W узлов коммутации пакетов, необходимо ввести ряд упрощающих предположений. Первое из них - предположение о независимости - позволяет устранить зависимость между временами обслуживания в каналах и состоит в том, что длина пакета, поступающего в i -й канал, выбирается независимо в соответствии с плотностью распределения:

$$f(x) = be^{-bx}$$

где b - средняя длина пакета, измеряемая в битах (байтах).

Процесс поступления пакетов в сеть является пуассоновским с параметром Λ_r (пакетов/с), где r - номер пары узел-источник – узел-адресат. Все пары упорядочены в соответствии с номерами $1, 2, \dots, R$. Маршрут пакетов r -го класса (передаваемых в r -й паре источник-адресат) определяется матрицей $\|P_{ij}(r)\|$, где $P_{ij}(r)$ - вероятность того, что пакет r -го класса, закончивший обслуживание в i -м канале, поступит потом в j -й канал ($i, j = \overline{1, M}$).

Различные способы задания матрицы $\|P_{ij}(r)\|$ определяют тип маршрутизации пакета в базовой сети. Например, при использовании постоянных виртуальных соединений и фиксированной маршрутизации соответствующие элементы $P_{ij}(r)$ принимают лишь два значения: 0 и 1. В рассматриваемой модели также предполагается, что объемы буферных накопителей не ограничены и подтверждение об успешной доставке пакета передается мгновенно.

Сделанные выше предположения позволяют полностью определить разомкнутую неоднородную сеть МО, моделирующую функционирование базовой сети передачи данных. В указанную сеть МО поступают r классов пуассоновских потоков пакетов с интенсивностями Λ_r ($r = \overline{1, R}$), маршрут каждого из которых характеризуется матрицей $\|P_{ij}(r)\|$. Функция распределения длительности обслуживания пакетов r -го класса в i -м центре сети МО, $i = \overline{1, M}$, который моделирует соответствующий канал передачи данных, является экспоненциальной с параметром $\mu_{ir} = C_i b_r$ (пакетов/с), где

C_i - пропускная способность i -го канала, измеряемая в бит/с, а b_r - средняя длина пакета r -го класса.

Интенсивность потока пакетов класса r , поступающих в i -й канал λ_{ir} удовлетворяет уравнению баланса потоков:

$$\lambda_{ir} = \Lambda_r \delta_{ir} + \sum_{j=1}^M \lambda_{jr} P_{ji}(r).$$

Здесь

$$\delta_{ir} = \begin{cases} 1, & \text{если поток } \Lambda_r \text{ поступает в } i\text{-й канал} \\ 0, & \text{в противном случае} \end{cases}$$

Потоки пакетов, поступающих в i -й канал λ_i , и общий поток Λ , поступающий в сеть извне, равны, соответственно:

$$\lambda_i = \sum_{r=1}^R \lambda_{ir}, \quad \Lambda = \sum_{r=1}^R \Lambda_r.$$

Обозначим также через ρ_{ir} загрузку i -го канала пакетами r -го класса и ρ_i общую загрузку канала i . Тогда:

$$\rho_{ir} = \frac{\lambda_{ir}}{C_i b_r} \quad \text{и} \quad \rho_i = \sum_{r=1}^R \rho_{ir}.$$

При расчёте описанной неоднородной открытой сети МО определяются вероятности стационарного состояния сети $P(n)$, где $n = \{n_1, n_2, \dots, n_M\}$, а значения n_i , $i = \overline{1, M}$ означают число пакетов, ожидающих передачи и передаваемых по i -му каналу. При этом, вероятность одного из n состояний равна:

$$P(\mathbf{n}) = \prod_{i=1}^M P_i(n_i).$$

Предполагая, что передача пакетов по каналу осуществляется в соответствии с дисциплиной FIFO, из последнего выражения можно определить среднее количество пакетов в i -м канале:

$$L_i = \frac{\rho_i}{1 - \rho_i}$$

и среднее число пакетов в сети в целом:

$$L = \sum_{i=1}^M \frac{\rho_i}{1 - \rho_i}.$$

В то же время известно, что в соответствии с формулой Литтла $L = \Lambda T$, где T - среднее время пребывания пакета в сети (задержка пакета).

Таким образом,

$$\Lambda T = \sum_{i=1}^M \frac{\rho_i}{1 - \rho_i} \text{ и } T = \frac{1}{\Lambda} \sum_{i=1}^M \frac{\lambda_i}{bC_i - \lambda_i}.$$

Данное выражение, впервые полученное Клейнроком, имеет важное значение для анализа и проектирования компьютерных сетей.

15. Управление потоками в сети с пакетной коммутацией

15.1. Методы управления потоками

Управление потоком предназначено для ограничения загрузки основных ресурсов сети (буферов УК и каналов связи) и согласования скорости передачи информации источником со скоростью приема адресатом. Являясь важнейшим компонентом сети пакетной коммутации, управление потоком реализуется на разных уровнях протоколов: оно может осуществляться между соседними УК базовой сети; между УК источником и УК адресатом базовой сети; между парами, обменивающимися информацией, и т.д. Выполнение функций согласования скоростей и надежной передачи информации на всех уровнях протоколов базируется на механизмах квитирования и концепции окна. Указанные механизмы служат также эффективным средством борьбы с перегрузкой ресурсов сети.

Отсутствие управления использованием ограниченных ресурсов сети при чрезмерном увеличении потока требований от одного или группы абонентов может привести к резкому увеличению времени задержки и (или) падению производительности сети вплоть до образования полностью блокированных участков, в которых передача данных оказывается невозможной. Причина появления блокированных участков и падения производительности сети при перегрузках могут быть различными. Ограничимся иллюстрацией этого явления лишь на двух простых примерах. На рис. 15.1 показаны два узла коммутации пакетов, связанные каналом передачи данных. Если буферная память узла A занята пакетами для узла B , а память узла B заполнена пакетами, предназначенными узлу A , то передача между ними невозможна из-за от-

сутствия свободных буферов и указанный участок оказывается полностью заблокированным.

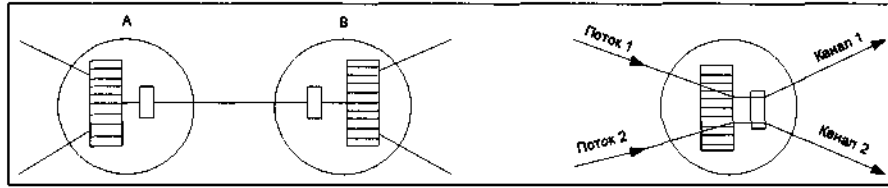


Рис. 15.1.

Рис. 15.2

На рис. 15.2 изображен УК, в который поступают два потока пакетов. Пакеты каждого потока передаются по своему, исходящему из узла каналу. Если все буфера УК заняты пакетами одного из потоков, например первого, то передача пакетов по второму каналу оказывается невозможной. Это так же, как и в первом случае снижает производительность УК и соответственно производительность сети в целом.

В общем случае при перегрузке сети значительно возрастает вероятность того, что пакеты, поступающие в очередной УК, не застают в нем свободных буферов. Эти пакеты в зависимости от используемого в сети протокола передаются повторно либо соседним УК, либо УК источником. Чем сильнее перегрузка, тем большая часть ресурсов сети занята выполнением повторных передач и, следовательно, тем ниже производительность сети, не защищенной от перегрузки (рис. 15.3).

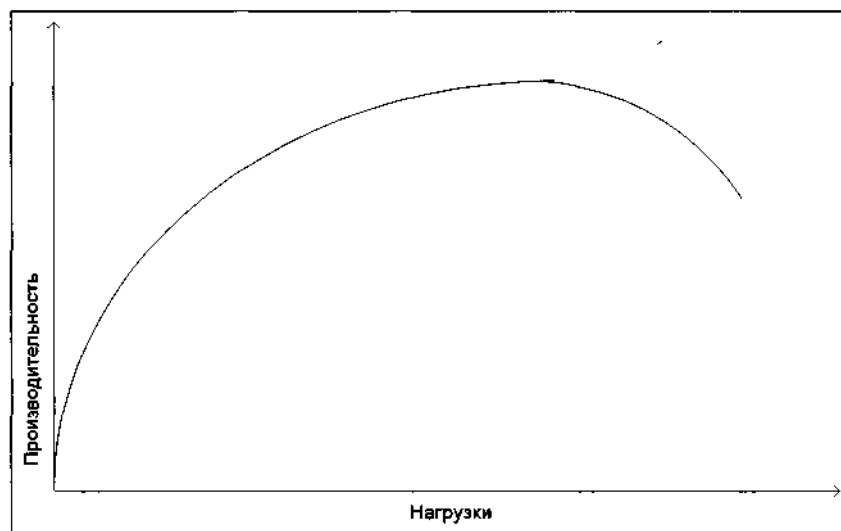


Рис. 15.3. Снижение производительности сети при перегрузках

Системы сетевого управления должны постоянно контролировать состояние сети и предотвращать возникновение опасных перегрузок. Чаще всего, в качестве основных симптомов сетевых перегрузок рассматриваются следующие ситуации:

- максимальная загруженность магистральных каналов превышает 70%;
- потеря пакетов (кадров, ячеек) превышает 2-3%;

- время задержки пакетов в сети увеличивается на 10%;
- значительное возрастание числа несовпадений контрольной суммы;
- приложения постоянно сталкиваются с тайм-аутами;
- биения PVC, т.е. ситуация, когда проложенное по сети виртуальное соединение (PVC), в силу каких-либо изменений в сети (отказы, очереди) меняет маршрут соединения, и когда эти изменения принимают непрерывный характер.

Защита от перегрузки достигается при управлении входным потоком (ограничении потока) в узлах базовой сети передачи данных. При этом различают методы глобального и локального управления нагрузкой. Глобальное управление предполагает либо ограничение общего количества пакетов, передаваемых в базовой сети (изаритмическое управление), либо ограничение числа пакетов в каждом виртуальном соединении источник-адресат или в группах виртуальных соединений (межконцевое управление), либо использование комбинации этих методов (двухуровневое управление). В отличие от глобального, локальное управление не требует информации о количестве передаваемых по сети пакетов и ограничивает поток в УК на основе локальной информации, касающейся только данного узла.

Изаритмическое управление реализуется с помощью использования в сети ограниченного числа служебных пакетов, содержащих разрешения на передачу пакетов данных. Возможно несколько способов получения разрешений. В первом случае пакет, поступивший в узел-адресат, освобождает принадлежащее ему разрешение и оно захватывается пакетом, ожидающим передачи. Если в узле отсутствуют пакеты, ожидающие передачи, то разрешение начинает циркулировать по сети случайным образом и захватывается в первом же узле, где имеются пакеты, ожидающие передачи. Во втором случае каждый узел содержит ограниченный пул разрешений, что позволяет уменьшить задержку пакетов из-за ожидания разрешений, циркулирующих по сети случайным образом. Возможна и комбинация этих подходов. Недостаток описанного метода глобального управления состоит в отсутствии удовлетворительной маршрутизации разрешений и трудностях практической реализации.

Реализация разнообразных способов межконцевого управления базируется на механизме окна, ограничивающем число пакетов в каждом виртуальном соединении (группа виртуальных соединений). Размер окна N_i для i -го виртуального соединения выбирается заранее и определяет число пакетов, которые могут быть переданы из узла-источника без подтверждения о правильном приеме узлом-адресатом. Счетчик числа пакетов в i -м виртуальном соединении первоначально устанавливается в состояние N_i . Каждый переданный и неподтвержденный пакет уменьшает состояние счетчика на единицу. Если состояние счетчика равно нулю, то новые пакеты в данное виртуальное соединение не передаются (поток пакетов от источника приостанавливается). Каждый правильно принятый в адресате пакет подтверждается индивидуально, причем прием квитанции АСК в узле-источнике увеличивает

состояние счетчика N_i на единицу. При изменении состояния счетчика с 0 на 1 источник потока пакетов вновь переводится в активное состояние.

Модель сети МО для исследования характеристик наиболее общего двухуровневого глобального управления, объединяющего изаритмическое и межконцевое управление, рассматривается ниже.

Наряду с описанным выше механизмом глобального управления в узлах сети часто применяют локальное управление, в соответствии с которым ограничивается либо общее число пакетов в УК, либо количество буферов, выделяемое разным классам пакетов. Принадлежность входящих в УК пакетов к разным классам может определяться по разным признакам: по числу пройденных пакетом транзитных участков, по принадлежности к виртуальным соединениям, по выходящим из узла направлениям и т. д. Ограниченная равнодоступная память УК при перегрузках может приводить к существенному падению пропускной способности узла за счет захвата памяти более интенсивными потоками отдельных классов и полной блокировки других. Локальное управление, ограничивающее число буферов, доступных «энергичным» потокам, часто называют управлением буферами. Такое управление наиболее типично для перегрузки, вызванной резким увеличением интенсивности потока пакетов в одно или несколько выходящих направлений узла.

Различают динамическое и статическое управление распределением буферов. Динамическое управление является наиболее общей стратегией управления буферами и состоит в принятии решения о вводе пакета в буферную память в момент его поступления в зависимости от состояния УК. Однако широкое использование этого управления ограничено трудностями практической реализации. Частыми случаями динамического управления являются легко реализуемые на практике различные статические механизмы распределения буферов: распределение без ограничений (CS) - любому входящему сообщению предоставляется любой свободный буфер; фиксированное распределение (CP) - за каждым классом пакетов закрепляется фиксированное число буферов; распределение с ограничением максимального числа буферов для различных классов (SMXQ); распределение с гарантированным минимумом буферов, закрепленным за каждым классом пакетов (SMA); распределение, являющееся объединением двух предыдущих методов локального управления (SMQMA).

15.2. Сетевая модель глобального управления

Рассмотрим базовую сеть передачи данных, включающую W узлов коммутации пакетов и M каналов связи. Очевидно, что общее количество виртуальных соединений (и соответственно число классов пакетов) равно

$$R = W(W - 1).$$

Пусть N_r , $r = \overline{1, R}$ - размер окна r -го виртуального соединения; Λ_r - интенсивность пуассоновского потока пакетов в r -е виртуальное соедине-

ние. Скорость поступления пакетов в r -е виртуальное соединение зависит от числа $0 < R_r < N_r$ не заквитированных пакетов, находящихся в этом соединении; если $R_r = N_r$, то очередной пакет r -го класса не принимается в сеть.

Таким образом,

$$\Lambda_j(N_r) = \begin{cases} \Lambda_r, & \text{при } R_r < N_r, \\ 0, & \text{при } R_r = N_r, \end{cases}$$

и, следовательно, в качестве модели рассматриваемой базовой сети с межконцевым механизмом управления потоком можно использовать замкнутую неоднородную сеть МО с R классами пакетов.

Предположим дополнительно, что количество буферов в узлах не ограничено, времена передачи пакетов по R -му каналу $R = \overline{1, M}$ независимы и распределены по экспоненциальному закону со средним значением $1/(bC_r)$

и маршрут пакетов r -го класса определяется матрицей $\|P_{ij}(r)\|$, $r = \overline{1, R}$. Тогда полученная замкнутая неоднородная сеть МО является локально-сбалансированной и для расчета ее характеристик могут быть использованы приближенные методы, описанные выше.

Размер окна N_r должен выбираться достаточно большим, так чтобы при флуктуациях потока в нормальном режиме не снижалась производительность (отсутствовала блокировка) и в то же время виртуальное соединение было надежно защищено от чрезмерного увеличения в нем потока. Если перегрузке подвержены все или большинство виртуальных соединений (общая перегрузка сети), то наряду с ограничениями N_r (ограничения первого уровня) целесообразно использование дополнительного ограничения N_0 (ограничение второго уровня) на общее число пакетов в сети, такого, что $N_0 < \sum_{i=1}^R N_r$.

Рассмотрим более подробно сетевую модель описанного выше механизма двухуровневого глобального управления. Предположим, что множество виртуальных сообщений разбито на $D \leq R$ непересекающихся групп. Каждая j -я группа ($j = \overline{1, D}$), обозначаемая Ω_j , может представлять собой, например, множество виртуальных соединений, исходящих из j -го узла коммутации пакетов. Предположим также, что на первом уровне введены ограничения N_j' на количество пакетов, которые могут передаваться в j -й группе. Очевидно, что при $D = R$ каждая группа содержит одно виртуальное соединение и $N_j' = N_j$.

Состояние сети МО, моделирующей описанное двухуровневое управление, определяется вектором $S = \{S_1, S_2, \dots, S_M\}$, где $S_R = \{S_{R1}, \dots, S_{RR}\}$ и S_{Rr} - число пакетов r -го класса (r -го виртуального соединения) в R -м канале. Допустимые состояния удовлетворяют условиям

$$\sum_{R=1}^M \sum_{r=1}^R S_{Rr} \leq N_0 \text{ и } \sum_{R=1}^M \sum_{r \in \Omega_j} n_{Rr} \leq N'_j,$$

где $n_{Rr} = 0$, если маршрут пакетов класса r не проходит через R -й канал (здесь для простоты предполагается, что маршрутизация является фиксированной). Сохраняя все предположения и введенные обозначения, легко видеть, что стационарные вероятности состояний сети имеют вид

$$P(S) = G^{-1} \prod_{R=1}^M S_R! \prod_{r=1}^R \frac{1}{n_{Rr}!} \left[\frac{\lambda_{Rr}}{bC_R} \right]^{n_{Rr}},$$

где $S_R = \sum_{r=1}^R S_{Rr}$; $\Lambda_{Rr} = \Lambda_r$ если маршрут пакетов r -го класса проходит через R -й канал, и $\Lambda_{Rr} = 0$ в противном случае.

Используя соответствующие формулы, можно легко определить основные характеристики базовой сети, в которой реализовано описанное выше двухуровневое управление потоками.

16. Расчёт числа каналов в сотовых сетях

Расчёт числа каналов в сотовых сетях проведём на примере анализа доступа мобильных станций (МС) к базовой станции (БС) в сети стандарта GSM для речевой информации. Стандартом GSM предусматривается работа речевого тракта в режиме коммутации каналов под управлением системы сигнализации SS-7 (Пакетные режимы GPRS и EDGE в данном случае не рассматриваются).

В стандарте GSM используется узкополосный многостанционный доступ с частотно-временным разделением каналов (TDMA). В структуре TDMA кадра GSM/900 содержится 8 временных позиций (окон) на каждой из 124-х несущих. Каждой MS для передачи речи предоставляется одно из окон на частоте передачи в полосе 890-915 МГц (линия “вверх”) и одно из окон на частоте приема в полосе 935-960 МГц (линия “вниз”). Каждая полоса разделена на 124 частотных канала с разносом в 200 кГц. Таким образом, физический канал в стандарте GSM представляет собой комбинацию временного и частотного разделения сигналов (Всего - $8 \times 124 = 992$ физических канала). Длина временного окна составляет 0.577 мс. Длина кадра равна $8 \times 0.577 = 4.616$ мс. При длительности одного бита равной 3.69 мкс, что соответ-

ствуется канальной скоростью в 271 кбит/с, информационная скорость в каждом из 8-и временных каналов составляет 24.7 кбит/с (суммарная информационная скорость в одном частотном канале составляет $8 \times 24.7 = 197.6$ кбит/с). Снижение скорости связано с тем, что в структуре окна из 156.25 бит информационными являются только 114 бит.

Рассматривая отдельную соту с одной БС и множеством МС как систему массового обслуживания, определим необходимое число рабочих каналов в БС при заданной вероятности потери вызова из-за их занятости. После этого можно будет определить необходимое число частотных каналов в БС.

Выбор числа рабочих частот в каждой соте, а также числа рабочих и управляющих каналов производится в зависимости от нагрузки, создаваемой абонентами данной соты. Фирма Motorola предлагает для определения чисел каналов рассчитывать общую нагрузку на БС в предположении, что нагрузка от одного абонента (удельная нагрузка) составляет $y = 0.025$ Эрл (Это соответствует типичному потоку вызовов с интенсивностью $\lambda = 1$ выз/час при средней продолжительности разговора – $t = 90$ с).

Табл. 16.1 содержит сведения, необходимые для определения числа каналов, а следовательно, и числа частот на БС, в зависимости от числа абонентов (нагрузки), обслуживаемых этой станцией. В стандарте GSM в каждом частотном канале имеется 8 временных окон, из которых небольшая часть выделяется под каналы управления (ССН), а основная часть - под рабочие каналы (ТСН). Например, при 4-х частотных каналах из 32-х временных окон 30 отведены под рабочие каналы, а два оставшихся окна используются под управляющие каналы.

Таблица 16.1. Зависимость чисел каналов различных типов от нагрузки

Число частотных каналов	Число рабочих Каналов	Нагрузка в Эрлантах (У)	
		$P_{отк} = 0.01$	$P_{отк} = 0.02$
1	7	2.5	2.94
2	14	7.4	8.20
3	22	13.7	14.9
4	30	20.3	21.9
5	38	27.3	29.2
6	45	33.4	35.6
7	53	40.6	43.1
8	61	47.9	50.6
9	69	55.2	58.2
10	77	62.6	65.8

Общая телефонная нагрузка на БС в час наибольшей нагрузки (ЧНН) определяется как:

$$Y = N y = N \lambda t, \quad (16.1)$$

где N – число абонентов, обслуживаемых данной БС в ЧНН, а y – удельная нагрузка от одной МС. Величина обслуживаемой нагрузки зависит

как от числа каналов, так и от требуемого качества обслуживания, которое в данном случае задаётся в виде допустимой вероятности потерь (вероятности отказа в предоставлении канала - $P_{\text{отк}}$).

Для расчёта числа частотных каналов при известном числе абонентов необходимо по формуле (16.1) найти нагрузку в Эрлангах и по табл. 16.1 подобрать соответствующую строку для требуемого качества обслуживания.

Например, при числе абонентов $N = 300$ по (16.1) получим $Y = 7.5$ Эрл. Тогда, при требуемом качестве обслуживания – не более 1% потерь ($P_{\text{отк}} = 0.01$), придётся выделить 3 частотных канала (22 канала ТСН), так как 2 частотных канала при этих требованиях смогут обслужить только 7.4 Эрл нагрузки. Если же смягчить требования по допустимым потерям, например, увеличить допустимые потери до $P_{\text{отк}} = 0.02$, то можно обойтись 2-я частотными каналами (14 каналов ТСН).

Расчёт пропускной способности сотовых систем для стандартов 3G и 4G, которые предназначены в большей степени для передачи данных в пакетном режиме, должен учитывать целый ряд таких чисто радиотехнических аспектов как распространение сигналов в сложном рельефе, специфику технологий QAM, OFDMA и MIMO. Эти особенности выводят данную задачу из области задач, решаемых теорией телетрафика, являющейся, по существу, распространением теории массового обслуживания на телекоммуникационные системы.

17. Исследование телекоммуникационных систем и сетей методом имитационного моделирования

Рассмотренные выше аналитические методы исследования телекоммуникационных сетей и их элементов очень удобны, доступны, не требуют больших финансовых и временных затрат, но обладают одним очень существенным недостатком, зачастую делающим их бесполезными. Причина их непригодности для многих реальных задач кроется в том, что получение аналитических решений связано с принятием целого ряда допущений и прежде всего предположение о пуассоновском характере входного потока и экспоненциальном распределении длительности обслуживания, а также отсутствием корреляции между процессами в различных узлах и каналах. Это позволяет рассматривать анализируемую систему как систему без последствия, что значительно упрощает математические выкладки. Особенно трудны для аналитических решений вопросы, связанные с анализом, а тем более с синтезом конфигурации сети.

В связи с этим весьма актуальными становятся задачи исследования телекоммуникационных систем и сетей методом имитационного моделирования. Этот метод предусматривает создание компьютерной программы, имитирующей последовательно во времени все процессы, происходящие в исследуемой сети. При этом нужно иметь в виду, что к понятию “все процессы” нужно подходить с большой осторожностью. Например, моделируя сеть с

коммутацией пакетов, можно в качестве моделируемых элементов рассматривать только узлы (маршрутизаторы, коммутаторы) и каналы. При этом имитации подлежат времена задержек пакетов в узлах и каналах в соответствии с тем или иным распределением длительностей этих задержек.

Однако, при желании получить более точные результаты, можно каждый узел рассматривать как совокупность нескольких портов (контроллеров), центрального микропроцессора, общей шины, памяти (буферы ожидания) и др. Желая получить ещё более точные результаты, можно каждый контроллер представлять совокупностью микропроцессора, памяти, интерфейсов и др., а работу каждого микропроцессора, в том числе и центрального, имитировать работой устройства управления, АЛУ и процессорной памяти.

При большом желании можно имитировать работу операционной системы и распространение электромагнитной волны по линии связи. Но! Чем больше мы будем проникать вглубь происходящих процессов, детализировать их и имитировать эти процессы с помощью моделирующей программы, тем более громоздкими будут эти программы, тем больше дней, месяцев, а может быть и лет труда высококвалифицированных программистов будет потрачено на создание такой программы.

Другим фактором, ограничивающим желание исследователя чрезмерно углубляться в детали функционирования моделируемой системы, является продолжительность прогона, которая для набора достаточной статистики тоже может продолжаться дни и месяцы. Добавим к этому, что имитационное моделирование не позволяет по набору исходных данных (расположение абонентов, тяготение между ними для различных видов трафика, требования по надёжности и многое другое) синтезировать нужные сети. Как правило, процедура синтеза сети заменяется анализом большого количества (20 – 40) различных вариантов проектируемой сети, предоставляемых группой экспертов, и выбором наилучшего варианта по результатам анализа.

В качестве примера, в приложении представлены две имитационные модели, используемые в МТУСИ СКФ при изучении таких дисциплин как «Сети связи» и «Мультисервисные сети связи»:

- Исследование пропускной способности сети с коммутацией пакетов;
- Моделирование мультисервисной сети связи.

Лабораторная работа № 1. Исследование пропускной способности сети с коммутацией пакетов

1. Цель работы. Изучить принципы организации информационного обмена в сети с коммутацией пакетов с дейтаграммным режимом. Исследовать зависимость качественных показателей сети от структуры сети и канальной емкости.

2. Краткие сведения об объекте моделирования

2.1. Общие принципы построения сети КП.

Сеть с коммутацией пакетов (СКП) представляет собой совокупность центров коммутации пакетов (ЦКП) и каналов передачи данных (КПД) их соединяющих.

Коммутация пакетов (КП) предусматривает разбивку длинных сообщений на отдельные порции - пакеты, снабжение каждого пакета адресом получателя и другими необходимыми для передачи по сети атрибутами и последовательную передачу пакетов от одного ЦКП к другому до ЦКП получателя. При этом сеть в соответствии с протоколами сетевого уровня (3-ий уровень ЭМВОС) выполняет процедуры маршрутизации, защиты от ошибок, организации очередей (буферизация в ЦКП), управления потоками и др.

Разбивка сообщений на пакеты выполняется на уровне транспортного канала (4-ый уровень ЭМВОС), который вместе с другими (нижележащими) уровнями реализуется в абонентском комплекте, например, с помощью сборщиков/разборщиков пакетов (packet assembler/disassembler - PAD).

Для каждого пакета в соответствии с его адресными признаками мы будем различать ЦКП-отправитель и ЦКП-получатель. Первый обеспечивает абоненту-отправителю точку входа в СКП, а второй - точку выхода к абоненту-получателю.

Длина поля данных в различных протоколах строго не оговаривается. Например, в сетях X.25 рекомендуется максимальная длина - 128 октетов, но допускаются значения от 16 до 1024 октетов (дискретно - в значениях 2^n).

Дейтаграммный режим передачи пакетов, рассматриваемый в данной работе, в отличие от режима виртуальных каналов, предполагает выдачу пакетов в сеть без предварительного установления соединения. При этом пакеты передаются независимо друг от друга, могут доводиться до ЦКП получателя по различным маршрутам и, в общем случае, могут поступать к абоненту (на 4-ый уровень) неупорядоченно, а также с потерями или дублями пакетов.

Одна из главных задач ЦКП - маршрутизация пакетов, т.е. выбор направления передачи. При децентрализованном алгоритме маршрутизации

каждый ЦКП решает эту задачу самостоятельно по заранее определенным маршрутам (статическая процедура) или на основании текущей сетевой информации о загрузке и исправности сети и ее элементов (динамическая процедура).

В любом случае каждый ЦКП содержит таблицу маршрутизации (статическую или динамически обновляемую), в которой каждому адресу получателя (или адресу сети, в которой находится получатель) соответствует одно из исходящих из данного ЦКП направлений передачи данных. Часто в таблицах маршрутизации указывают не одно, а два или три направления выдачи, упорядоченные в порядке предпочтительности.

Другая важная задача ЦКП - организация приема, хранения и выдачи пакетов в каналы связи. При этом очередность выдачи может отличаться от принципа FIFO (первым пришёл - первым обслужен) и учитывать различные признаки пакета (вид сообщения, категория срочности и др.). В случае занятости наиболее приоритетного направления, пакет может быть выдан в следующее по оптимальности направление, а при его отсутствии пакет устанавливается в очередь (хранится в буфере ожидания). Если пакет поступает в ЦКП в момент, когда заняты все каналы в требуемых направлениях и заполнен буфер ожидания, то он теряется (в соответствии с реализованным протоколом).

Очереди пакетов в ЦКП негативно влияют на главный показатель сети - время доведения (задержки) пакетов. Однако очереди значительно повышают коэффициент использования каналов и позволяют сократить потери пакетов в сети.

2.2.Исследуемые зависимости.

К числу главных показателей качества функционирования СКП относятся:

- время доведения пакетов от ЦКП-отправителя до ЦКП-получателя;
- вероятность потери пакетов в процессе передачи (уничтожение или искажение).

Характерной особенностью требований к времени доведения является то, что средние значения этого времени, которые, как правило, легко определяются аналитическими методами, не интересуют заказчика сети. Более важным является показатель вероятности превышения времени доведения определенного порога, задаваемого для каждой пары ЦКП, либо для сети в целом. Таким образом, перед аналитиками стоит задача в определении распределения $P(T)$ - вероятностей доведения пакета за время T .

Практически требование заказчика может задаваться как допустимая доля пакетов, доводимых за сверхдопустимое время $P_{\text{дов}}(T > T_{\text{доп}}) < R$. Например: $P_{\text{дов}}(T > 1\text{с}) < 0.01$, т.е. не менее 99% всех переданных пакетов должно быть доведено за время, не превышающее 1с.

Время доведения пакетов зависит от таких параметров сети, как:

- интенсивность и распределение входных потоков,

- конфигурация сети,
- числа каналов в ветвях и каналные скорости передачи данных,
- производительность ЦКП и длительность обработки пакетов в ЦКП,
- алгоритмы маршрутизации, управления потоками, защиты информации и др.

Входные потоки в СКП могут быть заданы в виде матрицы интенсивностей входных потоков Λ , элементы которой λ_{ij} определяют интенсивность потока от абонентов ЦКП- i , адресованного абонентам ЦКП- j .

Вероятность того, что отправителем случайно выбранного пакета является ЦКП- i , можно определить из матрицы Λ как:

$$P_i = \lambda_i / \Lambda_\Sigma \quad (1)$$

где λ_i - суммарный поток, исходящий из ЦКП-отправителя i , Λ_Σ - суммарный поток, исходящий из всех ЦКП-отправителей. Очевидно, что

$$\lambda_i = \sum_j \lambda_{ij} \quad \text{и} \quad \Lambda_\Sigma = \sum_i \lambda_i = \sum_{i,j} \lambda_{ij}$$

Вероятность того, что получателем случайно выбранного пакета, исходящего из ЦКП- i является ЦКП- j , определяется как:

$$P_{ij} = \lambda_{ij} / \lambda_i, \quad \sum_j P_{ij} = 1 \quad (2)$$

Кроме значений интенсивности, входные потоки характеризуются распределением интервалов между вызовами (может задаваться индивидуально для каждого потока, но, как правило, задается одинаковым для всей сети). То же самое можно сказать и о длине передаваемых пакетов - для упрощения анализа принимается, что средние значения и распределения длин пакетов одинаковы для любых пар абонентов.

Конфигурация сети и каналная емкость сети задаются матрицей каналов V , элементы которой V_{ij} определяют числа каналов, исходящих из ЦКП- i и входящих в ЦКП- j .

Необходимо отметить, что реальные сети с коммутацией пакетов создаются, как правило, из одноканальных ветвей различной пропускной способности.

3. Вопросы для домашней подготовки

3.1. Что такое метод коммутации пакетов?

3.2. Какие существуют режимы коммутации пакетов и в чем их различие?

3.3. Что является источником пакетов для СКП?

3.4. Чем регламентируется длина пакета?

3.5. Для чего создаются таблицы маршрутизации и какая информация в них содержится?

3.6. Как организуется выдача пакетов из ЦКП в каналы передачи данных?

3.7. На какие показатели сети КП влияют очереди к каналам ПД?

3.8. Какие основные показатели качества функционирования СКП вы знаете?

3.9. Как задаются требования ко времени доведения пакетов?

3.10. От каких параметров и как (качественно) зависит время доведения пакетов?

3.11. Как задаются интенсивности входных потоков пакетов?

3.12. Что означают и чему равны вероятности P_i и P_{ij} ?

3.13. Какими параметрами характеризуются длины пакетов?

3.14. Как задаются канальные ёмкости в ветвях сети?

4. Описание имитационной модели

4.1. Реализация основных элементов модели.

В представленном в пункте 4.2 листинге моделирующей программы за единицу модельного времени (е.м.в.) принят интервал в 10 мкс. Это означает, что время передачи пакета длиной, например, $L = 500$ байт по каналу со скоростью $C = 2$ Мбит/с составит:

$$T = 8L/C = 8 \cdot 500 / 2 \cdot 10^6 = 2000 \text{ мкс} = 200 \text{ е.м.в.}$$

Выбор ЦКП-отправителя выполняется в соответствии с распределением вероятностей (1), представленным в функции с меткой ОТР. А выбор ЦКП-получателя для данного ЦКП-отправителя - в соответствии с рядом распределений вероятностей (2), представленных в функциях с числовыми метками A_1, A_2, A_3, \dots , где последние цифры (1,2,3...) - номера ЦКП-отправителей.

В данной модели конфигурация сети в явном (графическом) виде не задается, но присутствует в плане распределение информации. В функциях с числовыми метками $A_{11}, A_{12}, A_{13} \dots$ каждому узлу получателю ставится в соответствии одна из ветвей, исходящих из ЦКП 1,2,3... . Реализована статическая процедура маршрутизации с одним путём к адресату из каждого узла.

Числа каналов в ветвях устанавливаются, как обычно в GPSS, оператором STORAGE, в котором идентификатор B12 соответствует ветви, исходящей из ЦКП-1 и входящей в ЦКП-2. В исходном варианте все ветви - одноканальные.

Канальные скорости задаются в виде числа единиц модельного времени, необходимого для передачи пакета средней длины (в рассматриваемом варианте эта длина равна 500 байтам). Распределение времени передачи по каналам в ветвях представлено функцией с меткой TPRD.

Для пакетов, ожидающих начала передачи в каналы связи, в центрах коммутации выделены отдельные буферы ожидания для каждой исходящей ветви. Распределение чисел мест ожидания по ветвям задаётся функцией с меткой DQ.

В исходном варианте все межузловые потоки λ_{ij} равны между собой. При составлении таблиц маршрутизации выбирались кратчайшие по числу ветвей пути, а при наличии нескольких равноценных путей в первую очередь (для некоторой определённости) выбирались наклонные (на графе) ветви, затем вертикальные и в последнюю очередь - горизонтальные ветви.

В результате эксперимента в стандартном отчете выдаются:

- распределение времени доведения пакетов по всем маршрутам,
- распределение времени доведения пакетов по одному из маршрутов, указанному в операторе с меткой MRSHR,
- числа заявок за прогон на выдачу пакетов по ветвям (матрица OBVETV) и по маршрутам (матрица OBMARSH).
- числа отказов (ликвидация пакета) из-за занятости каналов и мест ожидания с указанием номера ветви (матрица OTVETV) и номера маршрута (матрица OTMARSH),
- вероятности отказов по ветвям (матрица VVETV) и по маршрутам (матрица VMARSH).

Распределения времени представляются в табличной и графической форме.

4.2. Листинг моделирующей программы

```
; модель сети коммутации пакетов
; 1 единица модельного времени равна 10 мкс
vhdpot equ 5000 ;интенсивность входного потока (число пакетов в сек)
tpred equ 4000 ;допустимое время доведения пакетов в сети
dq function p$vetv,d24 ;распределение числа мест ожидания в ветвях сети
12,10/21,10/23,10/32,10/34,10/43,10/14,10/41,10/
56,10/65,10/67,10/76,10/78,10/87,10/58,10/85,10/
15,10/51,10/26,10/62,10/37,10/73,10/48,10/84,10
tprd function p$vetv,d24 ;пропускные способности каналов в ветвях сети
12,200/21,200/23,200/32,200/34,200/43,200/14,200/
41,200/56,200/65,200/67,200/76,200/78,200/87,200/
58,200/85,200/15,200/51,200/26,200/62,200/37,200/
73,200/48,200/84,200
aa1 equ 1 ;номера узлов
aa2 equ 2
aa3 equ 3
aa4 equ 4
aa5 equ 5
aa6 equ 6
aa7 equ 7
aa8 equ 8
aa11 equ 11 ;номера таблиц маршрутизации
aa12 equ 12
aa13 equ 13
aa14 equ 14
aa15 equ 15
aa16 equ 16
aa17 equ 17
aa18 equ 18
b12 equ 12 ;номера ветвей
b21 equ 21
b23 equ 23
```


b32	equ	32	
b34	equ	34	
b43	equ	43	
b14	equ	14	
b41	equ	41	
b56	equ	56	
b65	equ	65	;номера ветвей
b67	equ	67	
b76	equ	76	
b78	equ	78	
b87	equ	87	
b58	equ	58	
b85	equ	85	
b15	equ	15	
b51	equ	51	
b26	equ	26	;номера ветвей
b62	equ	62	
b37	equ	37	
b73	equ	73	
b48	equ	48	
b84	equ	84	
b12	storage	1	;числа
b21	storage	1	
b23	storage	1	;каналов
b32	storage	1	
b34	storage	1	;в ветвях
b43	storage	1	
b14	storage	1	;сети
b41	storage	1	
b56	storage	1	
b65	storage	1	
b67	storage	1	
b76	storage	1	
b78	storage	1	
b87	storage	1	
b58	storage	1	
b85	storage	1	
b15	storage	1	
b51	storage	1	
b26	storage	1	
b62	storage	1	
b37	storage	1	
b73	storage	1	
b48	storage	1	
b84	storage	1	
mrez	matrix	,5,1	
obmarsh	matrix	,8,8	
otmarsh	matrix	,8,8	
vmarsh	matrix	,8,8	
obvetv	matrix	,8,8	
otvetv	matrix	,8,8	
vvetv	matrix	,8,8	
tdov	table	m1,0,200,40	
tmar	table	m1,0,200,40	
otp	function	rn1,d8	
.125,1/.25,2/.375,3/.5,4/.625,5/.75,6/.875,7/1,8			
aa1	function	rn1,d7	;получатели

```

.143,2/.286,3/.429,4/.572,5/.715,6/.858,7/1,8
aa2  function  rn1,d7
.143,1/.286,3/.429,4/.572,5/.715,6/.858,7/1,8
aa3  function  rn1,d7
.143,1/.286,2/.429,4/.572,5/.715,6/.858,7/1,8
aa4  function  rn1,d7
.143,1/.286,2/.429,3/.572,5/.715,6/.858,7/1,8
aa5  function  rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,6/.858,7/1,8
aa6  function  rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,5/.858,7/1,8
aa7  function  rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,5/.858,6/1,8
aa8  function  rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,5/.858,6/1,7
aa11 function  p$pol,d7      ;таблица маршрутизации
2,12/3,14/4,14/5,15/6,15/7,15/8,15
aa12 function  p$pol,d7
1,21/3,23/4,23/5,26/6,26/7,26/8,26
aa13 function  p$pol,d7
1,32/2,32/4,34/5,37/6,37/7,37/8,37
aa14 function  p$pol,d7
1,41/2,41/3,43/5,48/6,48/7,48/8,48
aa15 function  p$pol,d7
1,51/2,51/3,51/4,51/6,56/7,58/8,58
aa16 function  p$pol,d7
1,62/2,62/3,62/4,62/5,65/7,67/8,67
aa17 function  p$pol,d7
1,73/2,73/3,73/4,73/5,76/6,76/8,78
aa18 function  p$pol,d7
1,84/2,84/3,84/4,84/5,85/6,85/7,87
mrshr equ      17
raspr function  rn1,c24
0,0/.1,.104/.2,.222/.3,.355/.4,.509/.5,.69/.6,.915/.7,1.2/.75,1.38/
.8,1.6/.84,1.85/.88,2.12/.9,2.3/.92,2.52/.94,2.81/.95,2.99/.96,3.2/
.97,3.5/.98,3.9/.99,4.6/.995,5.3/.998,6.2/.999,7/1,8
per  variable  100000/vhdpot
mr   variable  p$otpr#10+p$pol
pu   variable  p$pol+10
ku   variable  p$otpr+10
iii  variable  p$vetv/10
jj   variable  p$vetv@10+10
jjj  variable  p$vetv@10
oger variable  (qa$b12+qa$b14+qa$b15+qa$b73+qa$b76+qa$b78)/6
isp  variable  (sr$b12+sr$b14+sr$b15+sr$b73+sr$b76+sr$b78)/6000
vetv fvariable 1000#mx$otvetv(p$iii,p$jjj)/mx$obvetv(p$iii,p$jjj)
marsh fvariable 1000#mx$otmarsh(p$otpr,p$pol)/mx$obmarsh(p$otpr,p$pol)
ver  fvariable x$otkaz#1000/(n$fin+1)+0.5
ot   fvariable x$otkaz/n$fin
zz   fvariable x$zader/n$fin
;-----
generate v$per,fn$raspr
assign  otpr,fn$otpr
assign  pol,fn*otpr
assign  mar,v$mr
assign  ku,v$ku
cicl   assign  vetv,fn*ku

```

```

assign    iii,v$iii
assign    jjj,v$jjj
msavevalue obvetv+,v$iii,v$jjj,1
test ge   q*vetv,fn$ddq,im
gate snf  p$vetv,otk
im  queue p$vetv
enter    p$vetv
depart   p$vetv
advance  fn$tpred,fn$raspr
leave    p$vetv
assign    ku,v$jj
test e    v$pu,p$ku,cicl
tabulate  tdov
test e    p$mar,mrshr,aaa
tabulate  tmar
aaa  test g    m1,tpred,fin
savevalue zader+,1
transfer  ,fin
otk  savevalue otkaz+,1
msavevalue otmarsh+,p$otpr,p$pol,1
msavevalue otvetv+,v$iii,v$jjj,1
fin  savevalue verot,v$ver
msavevalue vvetv,v$iii,v$jjj,v$vetv
msavevalue obmarsh+,p$otpr,p$pol,1
msavevalue vmarsh,p$otpr,p$pol,v$marsh
msavevalue mrez,1,1,n$fin
msavevalue mrez,2,1,v$ot
msavevalue mrez,3,1,v$zz
msavevalue mrez,4,1,v$isp
msavevalue mrez,5,1,v$oger
terminate 1

```

5. Лабораторное задание

5.1. Изучить модель СКП, способы ввода исходных данных и считывания результатов эксперимента.

5.2. Измерить выходные параметры для заданных в работе исходных данных и конфигурации СКП (однородная структура типа “сплюсненный куб” - рис.П.1).

В результирующей матрице MREZ основные результаты прогона фиксируются в следующей последовательности:

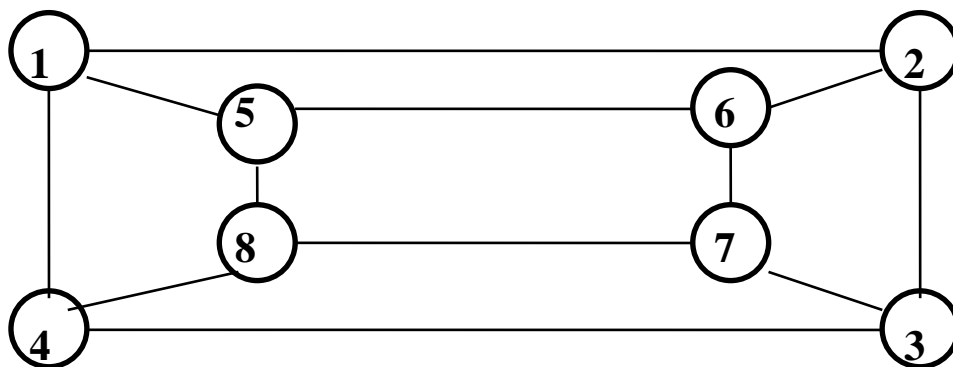


Рис. П.1. Схема моделируемой сети

- V – 1-я строка матрицы - число обработанных пакетов (доведённых и снятых с обслуживания). Оно равно числу транзактов, установленных в команде START;

- n - 2-я строка - число потерянных (не доведённых) пакетов, т.е. снятых с обслуживания из-за занятости всех каналов и мест ожидания в буфере в какой-либо из ветвей по маршруту передачи этого пакета;

- z - 3-я строка - число пакетов, доведённых до адресата за время, превышающее допустимое. Допустимое время доведения $T_{\text{доп}}$ задаётся оператором TPRED в единицах модельного времени. Например, оператор

TPRED EQU 5000

установит допустимое время в 0.05 с (Напомним, что 1 е.м.в.=10 мкс);

- k - 4-я строка - средний коэффициент использования (средняя загрузка) каналов по всей сети.

- q - 5-я строка - средняя длина очереди к каналам по всей сети;

5.3. Изменить исходные данные, установив входной поток Λ по номерам бригад в соответствии с табл. П.1. Входной поток задаётся оператором VHDPOT в числах пакетов в секунду.

Таблица П.1. Выбор входных параметров по номеру бригады

Номер бригады	1	2	3	4	5	6	7	8	9	10
Входной поток, тыс	4.8	4.9	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8

Снять зависимость показателей n, z, k и q от изменения входного потока Λ в точках Λ ; 1.4 Λ ; 1.7 Λ и 2 Λ . Результаты занести в табл. 4.2.

Таблица П.2. Результаты экспериментов по прогонам

№ эксперимента		1	2	3	4	5
Условия	Λ ,	5000					
	C,	2					
	Q	10					
Результаты	d	.013					
	z	.007					
	k	.728					
	q	1.724					

5.4. Для номинального входного потока Λ (табл. 4.1) снять зависимость основных показателей n, z, k и q от числа мест в буферах ожидания - Q. Значения Q задаются в операторе DQ FUNCTION P\$VETV,D24 для каждой из 24-х ветвей. В исходном варианте Q=10. Значения Q в прогонах устанавливать равными 0, 5, 10, 15 и 20.

5.5. Для номинального входного потока Λ (табл. 4.1) и значений Q

равных 10-и снять зависимость основных показателей n , z , k и q от скорости магистрального канала. В исходном варианте $C = 2$ Мбит/с. Значения C в прогонах устанавливать равными 1, 2 и 4 Мбит/с. Канальные скорости задаются функцией `TPRD FUNCTION P$VETV,D24`. В исходном варианте в этом операторе для каждой из 24-х ветвей установлено значение 200 е.м.в. (время передачи пакета длиной в 500 байт по каналу с полосой в 2 Мбит/с). Для скорости в 1 Мбит/с это время нужно увеличить вдвое, а для скорости в 4 Мбит/с – уменьшить вдвое.

Подобрать опытным путём минимально необходимую канальную скорость для выполнения следующих условий (отдельно по каждому условию):

- характеристика потерь, как максимально допустимая доля потерянных пакетов: $P_{\text{пот}} < 0.05$. Определяется как значение d в MREZ;
- характеристика доведения, как допустимая доля пакетов, доводимых за сверхдопустимое время: $P_{\text{дов}}(T > 0.1\text{с}) < 0.05$. Определяется как значение z в MREZ.

Зафиксировать в отчёте канальную скорость, достаточную для выполнения обоих условий.

5.6. Перестроить параметры сети следующим образом: заменить одноканальные ветви с пропускной способностью $C = 2$ Мбит/с на многоканальные ветви с 32-я каналами в каждой ветви и скоростью в каждом канале, равной $C_1 = 64$ кбит/с. Числа каналов изменяются в операторе `STORAGE`, а канальные скорости задаются функцией `TPRD FUNCTION P$VETV,D24`. Исходное значение 200 е.м.в. необходимо заменить для каждой ветви на значение 6250 е.м.в. (время передачи пакета длиной в 500 байт по каналу с полосой 64 кбит/с). Таким образом, примерно сохраняются пропускные способности ветвей.

Провести прогон для номинального входного потока Λ (табл. 4.1) и $Q = 10$. Полученные результаты (n , z , k и q) и гистограммы сопоставить с результатами по п. 5.3 для тех же значений Λ и Q .

6. Отчёт по работе

6.1. Описать работу модели СКП и перечень вводимых в модель и измеряемых параметров.

6.2. Построить графики зависимостей и гистограммы по пп. 5.3, 5.4 и 5.5.

6.3. Объяснить результаты измерений по п. 5.6.

7. Контрольные вопросы

7.1. Как выбирается единица модельного времени?

7.2. Как формируется суммарный входной поток пакетов?

7.3. Опишите процедуру формирования частных потоков (от конкретного ЦКП-отправителя к конкретному ЦКП-получателю).

7.4. Опишите процедуру формирования таблицы маршрутизации.

7.5. Объясните зависимость основных показателей сети от объема буферов ожидания и от интенсивности входного потока.

Лабораторная работа №2. Моделирование мультисервисной сети связи

1. Цель работы:

Изучить основные принципы построения имитационных моделей сложных мультисервисных сетей связи.

Ознакомиться с основными приёмами анализа сложных сетей связи на имитационной модели.

2. Описание моделируемой системы

Моделируется мультисервисная сеть связи городского масштаба, состоящая из 8-и составных узлов (СУ), связанных в кольцевую структуру оптоволоконными линиями по технологии SDH (рис. П.2).

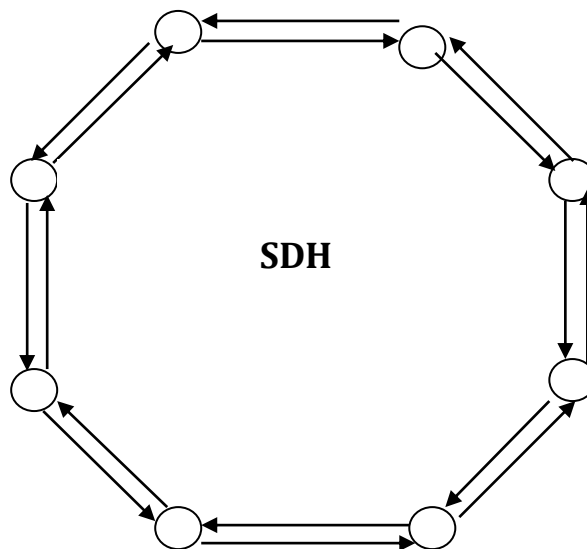


Рис.П.2. Схема мультисервисной сети.

Каждый из 8-и СУ (рис. П.3) является совокупностью узла абонентского доступа, узла коммутации вторичной сети и мультиплексора первичной сети (мультиплексор SDH).

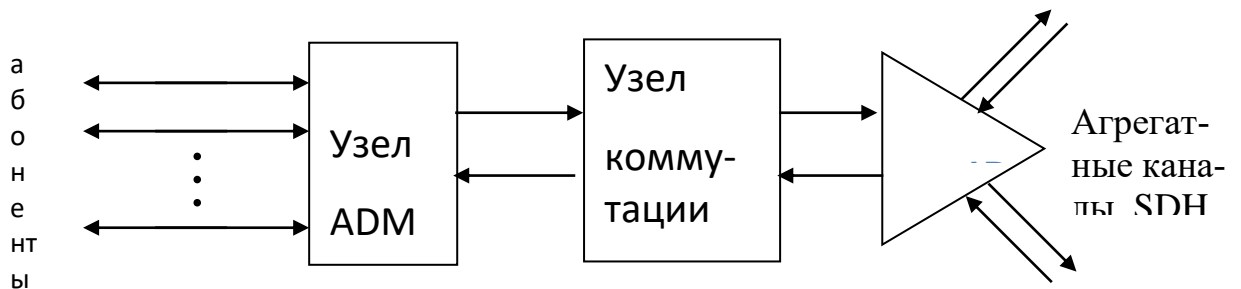


Рис. П.3. ADM – мультиплексор ввода-вывода.

Имитируемая система не предполагает реализацию какой-либо конкретной технологии абонентского доступа или коммутации вторичной сети.

В части абонентского доступа это могут быть любые технологии поддерживающие потоки аудио+видео+данные (триада A+V+D) или любую часть из них. Например, xDSL, ISDN, IP-пакеты, FR, ATM и др.

В части узла коммутации вторичной сети имитируемая система реализует режим работы, близкий к режиму быстрой коммутации пакетов (БКК), например, по технологии DTP.

3. Описание имитационной модели

3.1. Генерация входных потоков

Модель имитирует поступление потоков трех видов:

- Телефонные вызовы с устанавливаемыми интервалами между вызовами суммарного потока вызовов в сеть – $t_{и}^A$ и продолжительностью обслуживания (разговора) – t_o^A ;
- Вызовы видеоконференцсвязи (ВКС) с интервалами – $t_{и}^V$ и продолжительностью обслуживания – t_o^V ;
- Поток файлов с интервалами между их поступлениями в сеть – $t_{и}^D$ и объёмом файлов – Q^D , определяемом функцией с именем `dlina`.

Интенсивности каждого из входных потоков и длительности обслуживания устанавливаются независимо (3 отдельных оператора `generate`).

Распределение общих потоков между восемью узлами отправителя задаётся функцией `OTPR`. В исходном варианте это распределение равномерное, т.е. любой вызов любого вида может с вероятностью 0,125 (т.е. 1/8) поступить в один из 8-и узлов.

Восемь функций с метками $A1, A2, \dots, A8$ определяют распределение вызовов от каждого узла отправителя к семи узлам получателя. Эти распределения одинаковы для всех видов трафика.

В исходном варианте это распределение равномерное, т.е. получателем любого вызова любого вида может с вероятностью 0,143 (т.е. 1/7) стать один из семи узлов.

Вызовам (транзактам) присваиваются приоритеты:

- Телефонные – 2-й приоритет (высший);
- ВКС – 1-й приоритет;
- Данные (файлы) – 0-й приоритет (низший).

3.2. Обслуживание вызовов

Все вызовы обслуживаются в режиме виртуальных соединений с выделением из общего канального ресурса определённой квоты:

- для телефонных вызовов – 1 квота, за которую принята канальная скорость ОЦК – 64 кбит/с (стандарт DSO);
- для ВКС – 6 квот, что соответствует скорости 384 кбит/с;

- для передачи данных принят гибкий алгоритм, при котором для передачи файла выделяется только определённая часть из свободной в данный момент полосы канала. Эта часть может устанавливаться в параметре - doly.

Процедура установления соединения состоит в следующем:

- после определения номеров узлов отправителя и получателя программа определяет кратчайшее направление по кольцу от отправителя к получателю.

- по каждой ветви этого маршрута определяется возможность выделения необходимых квот полосы (для ТЛФ – 1 квота, Для ВКС – 6 квот, для данных – установленная в параметре doly доля свободных квот);

- соединения для ТЛФ и ВКС устанавливаются, если по всему маршруту имеются соответствующие свободные квоты (1 квота для ТЛФ и 6 квот для ВКС). В противном случае эти вызовы удаляются из системы и фиксируются как потерянные;

- при передаче файлов для каждой ветви в установленном маршруте определяются необходимые величины квот и минимальная из них выделяется для всего соединения. При невозможности выделения в какой-либо ветви по маршруту ни одной квоты передаче файла будет отказано и фиксируется потеря файла;

- дополнительное преимущество потоковым видам трафика (А и V) в данном алгоритме обеспечивается организацией так называемого резерва полосы в транспортном модуле (rezp). Смысл этого резерва состоит в том, что при определении возможности выделения квот для передачи файла данных (определение доли квот из числа свободных), резервная полоса исключается из рассмотрения. Это резерв для аудио и видео трафика.

Общее количество квот в ветви определяется следующим образом.

Для транспортного модуля STM-1 (скорость 155.52 Мбит/с) объём контейнера составляет 2430 байт, из которых для передачи информации могут использоваться 2349 байт. Остальные 81 байт – стандартом отведены под управление.

Контейнеры в SDH передаются с интервалами 125 мкс. Оцифрованная речь требует передачи по одному байту так же каждые 125 мкс (частота дискретизации 8 кГц). Следовательно одна квота соответствует 1 байту и общая полоса канала STM-1 составляет 2349 квот. Именно из этого числа будут заниматься квоты для всех видов информации в имитационной модели.

Моделирующая программа представлена ниже.

*		1 единица модельного времени = 0.1 с	
a	equ	1	;интервалы между тлф-вызовами(выз/с)
vv	equ	2	;интервалы между вкс-вызовами (выз/с)
d	equ	1	;интервалы между поступлением файлов (ф/с)
doly	equ	0.2	;доля свободной полосы для передачи файла
rezp	equ	150	;резервная полоса для аудио и видео трафика
asr	equ	1800	;средняя длит. разговора
vsr	equ	3000	;средняя длит. видео-конференции
tpred	equ	20	;предельно допустимое время задержки файлов
adlit	fvariable	int(asr#fn\$raspr)	;длительности а-обслуживания
vdlit	fvariable	int(vsr#fn\$raspr)	;длительности v-обслуживания


```

nkvot equ 2349
kvot variable int(r*vetv#doly) ;квоты для передачи данных
dlina function rn1,d10 ;длина передаваемого файла в кбайтах
0,1/.2,10/.3,50/.4,100/.5,200/.6,500/.7,700/.8,1000/.9,1500/1,3000
tim fvariable int(fn$dlina#1.25/p$kvota) ;продолжительность передачи файла
aver fvariable mx$mitog(2,1)/mx$mitog(1,1);вычисление вер.потери а-вызова
vver fvariable mx$mitog(4,1)/mx$mitog(3,1);вычисление вер.потери v-вызова
dver fvariable mx$mitog(6,1)/mx$mitog(5,1);выч.вер.отказа в передаче файла
zsr fvariable x$tzad/n$zz ;выч.среднего времени задержки файла в сети
zver fvariable n$zzz/n$zz ;выч.вер.превышения t-допустимого
obkrit fvariable ;обобщённый критерий
(au#mx$mitog(7,1)+vu#mx$mitog(8,1)+du#mx$mitog(9,1)+zu#mx$mitog(11,1))#10
gru variable x$gru@1000000+1 ;номера групп для link и unlink
au equ 50 ;условная стоимость потери а-вызова
vu equ 25 ;условная стоимость потери v-вызова
du equ 10 ;условная стоимость отказа в передаче файла
zu equ 7 ;условная стоимость задержки файла сверх t-допустимого
aa1 equ 1 ;nomera uzlov
aa2 equ 2
aa3 equ 3
aa4 equ 4
aa5 equ 5
aa6 equ 6
aa7 equ 7
aa8 equ 8
aa11 equ 11 ;nomera tablitz marshrutizayii
aa12 equ 12
aa13 equ 13
aa14 equ 14
aa15 equ 15
aa16 equ 16
aa17 equ 17
aa18 equ 18
b12 equ 12 ;nomera vetvey
b21 equ 21
b23 equ 23
b32 equ 32
b34 equ 34
b43 equ 43
b45 equ 45
b54 equ 54
b56 equ 56
b65 equ 65 ;nomera vetvey
b67 equ 67
b76 equ 76
b78 equ 78
b87 equ 87
b81 equ 81
b18 equ 18
mitog matrix ,13,1 ;матрица итоговых результатов
obmarsh matrix ,8,8
otmarsh matrix ,8,8
vmarsh matrix ,8,8
tdov table m1,0,10,50
tmar table m1,0,10,50
aotp function rn1,d8 ;выбор отправителя а
.125,1/.25,2/.375,3/.5,4/.625,5/.75,6/.875,7/1,8
votp function rn1,d8 ;выбор отправителя v
.125,1/.25,2/.375,3/.5,4/.625,5/.75,6/.875,7/1,8
dotp function rn1,d8 ;выбор отправителя d
.125,1/.25,2/.375,3/.5,4/.625,5/.75,6/.875,7/1,8
aa1 function rn1,d7 ;выбор получателя

```

```

.143,2/.286,3/.429,4/.572,5/.715,6/.858,7/1,8
aa2 function rn1,d7
.143,1/.286,3/.429,4/.572,5/.715,6/.858,7/1,8
aa3 function rn1,d7
.143,1/.286,2/.429,4/.572,5/.715,6/.858,7/1,8
aa4 function rn1,d7
.143,1/.286,2/.429,3/.572,5/.715,6/.858,7/1,8
aa5 function rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,6/.858,7/1,8
aa6 function rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,5/.858,7/1,8
aa7 function rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,5/.858,6/1,8
aa8 function rn1,d7
.143,1/.286,2/.429,3/.572,4/.715,5/.858,6/1,7
aa11 function p$pol,d7 ;таблицы маршрутизации
2,12/3,12/4,12/5,12/6,18/7,18/8,18
aa12 function p$pol,d7
1,21/3,23/4,23/5,23/6,23/7,21/8,21
aa13 function p$pol,d7
1,32/2,32/4,34/5,34/6,34/7,34/8,32
aa14 function p$pol,d7
1,43/2,43/3,43/5,45/6,45/7,45/8,45
aa15 function p$pol,d7
1,56/2,54/3,54/4,54/6,56/7,56/8,56
aa16 function p$pol,d7
1,67/2,67/3,65/4,65/5,65/7,67/8,67
aa17 function p$pol,d7
1,78/2,78/3,78/4,76/5,76/6,76/8,78
aa18 function p$pol,d7
1,81/2,81/3,81/4,81/5,87/6,87/7,87
b12 storage 2349 ;величины квот в ветвях
b21 storage 2349
b23 storage 2349
b32 storage 2349
b34 storage 2349
b43 storage 2349
b45 storage 2349
b54 storage 2349
b56 storage 2349
b65 storage 2349
b67 storage 2349
b76 storage 2349
b78 storage 2349
b87 storage 2349
b81 storage 2349
b18 storage 2349
raspr function rn1,c24 ;табличное экспоненциальное распределение
0,0/.1,.104/.2,.222/.3,.355/.4,.509/.5,.69/.6,.915/.7,1.2/.75,1.38/
.8,1.6/.84,1.85/.88,2.12/.9,2.3/.92,2.52/.94,2.81/.95,2.99/.96,3.2/
.97,3.5/.98,3.9/.99,4.6/.995,5.3/.998,6.2/.999,7/1,8
mrshr equ 17
per variable 100000/vhdpot
mr variable p$otpr#10+p$pol
pu variable p$pol+10
ku variable p$otpr+10
iii variable (p$vetv-p$vetv@10)/10
jj variable p$vetv@10+10
jjj variable p$vetv@10
isp variable (sr$b12+sr$b23+sr$b34+sr$b45+sr$b56+sr$b67+sr$b78+sr$b81)/8
marsh fvariable mx$otmarsh(p$otpr,p$pol)/mx$obmarsh(p$otpr,p$pol)
ver fvariable x$otkaz/(n$fin+1)+0.5

```

```

*-----
generate  A,fn$raspr,,,2      ;телефонный поток
assign    otrpr,fn$Aotpr      ;установка отправителя
assign    pol,fn*otpr         ;установка получателя
msavevalue mitog+,1,1,1
assign    ku,v$ku
aaa  assign    vetv,fn*ku
gate snf    p$vetv,Aotk
assign    ku,v$jj
test e     v$pu,p$ku,aaa
assign    kvota,1
assign    ttt,v$Adlit
transfer   ,obshee

*-----
generate  VV,fn$raspr,,,1     ;видеоконференцсвязь
assign    otrpr,fn$Votpr
assign    pol,fn*otpr
msavevalue mitog+,3,1,1
assign    ku,v$ku
vvv  assign    vetv,fn*ku
test ge    R*vetv,6,Votk
assign    ku,v$jj
test e     v$pu,p$ku,vvv
assign    kvota,6
assign    ttt,v$Vdlit
transfer   ,obshee

*-----
generate  D,fn$raspr          ;поток данных
assign    otrpr,fn$Dotp
assign    pol,fn*otpr
msavevalue mitog+,5,1,1
assign    ku,v$ku
savevalue delta,Nkvot
ddd  assign    vetv,fn*ku
test ge    R*vetv,rezp,Dotk    ;резерв полосы для А и V трафиков
test ge    delta,R*vetv,met1
savevalue  delta,R*vetv
assign    kvota,v$kvot
met1  test g    p$kvota,0,Dotk
assign    ku,v$jj
test e     v$pu,p$ku,ddd
assign    ttt,v$tim

*-----
obshee assign    ku,v$ku
savevalue  gru,v$gru
assign     gru,x$gru
assign     mar,v$mr
cicl  assign    vetv,fn*ku      ;выбор ветви по таблице маршрутизации
assign    iii,v$iii            ;исходящий узел ветви
assign    jjj,v$jjj            ;входящий узел ветви
split     1,met
gate snf    p$vetv,Dotk
enter      p$vetv,p$kvota
link       gru,fifo
met  assign    ku,v$jj
test e     p$ku,v$pu,cicl
advance    p$ttt
unlink     gru,un,all,gru,p$gru
test e     0,pr,fin
savevalue  Tzad+,m1
test g     m1,Tpred,zz
zzz  savevalue chzad+,1

```

```

*zz  tabulate  tdov
zz   test e    p$mar,mrshr,fin
*    tabulate  tmar
      transfer ,fin
Aotk  msavevalue Mitog+,2,1,1
      transfer ,otk
Votk  msavevalue Mitog+,4,1,1
      transfer ,otk
Dotk  msavevalue Mitog+,6,1,1
otk   savevalue  otkaz+,1
*     MSAVEVALUE OTMARSH+,P$OTPR,P$POL,1
*fin  SAVEVALUE  VEROT,V$VER
*     MSAVEVALUE OBMARSH+,P$OTPR,P$POL,1
*     MSAVEVALUE VMARSH,P$OTPR,P$POL,V$MARSH
fin   test e    tg1,1,final      ;транзакт последний?
      msavevalue Mitog,7,1,v$Aver
      msavevalue Mitog,8,1,v$Vver
      msavevalue Mitog,9,1,v$Dver
      msavevalue Mitog,10,1,v$Zsr
      msavevalue Mitog,11,1,v$Zver
      msavevalue Mitog,12,1,v$obkrit
      msavevalue Mitog,13,1,N$fin
final terminate 1
un    leave     p$vetv,p$kvota
      terminate

```

3.3. Регистрация событий в модели

Кроме результатов, помещаемых в стандартном отчёте, в данной модели снимаются следующие характеристики:

а). Матрица `mitog` в виде вектора столбца:

- 1-я строка – общее число ТЛФ вызовов;
- 2-я строка – число потерянных ТЛФ вызовов;
- 3-я строка – общее число ВКС вызовов;
- 4-я строка – число потерянных ВКС вызовов;
- 5-я строка – общее число инициированных файлов ПД;
- 6-я строка – число потерянных файлов ПД;
- 7-я строка – доля потерянных ТЛФ вызовов ($P_{\text{пот.ТЛФ}}$);
- 8-я строка – доля потерянных ВКС вызовов ($P_{\text{пот.ВКС}}$);
- 9-я строка – доля потерянных файлов ПД ($P_{\text{пот.ПД}}$);
- 10-я строка – доля файлов, задержанных сверх допустимого времени

(α). Допустимое время задержки устанавливается в параметре T_{pred} .

- 11-я строка – среднее время задержки файлов $T_{\text{ср}}$;
- 12-я строка – обобщённый критерий потерь, подсчитываемый как сумма взвешенных потерь

$$K = P_{\text{тлф}} * V_{\text{тлф}} + P_{\text{вкс}} * V_{\text{вкс}} + P_D * V_D + P_Z * V_Z$$

где P_i – доли соответствующих потерь по результатам прогона;

V_i – вес соответствующих потерь в обобщённом критерии – устанавливаются в параметрах A_u , V_u , D_u и Z_u .

• 13-я строка – общее число пропущенных через модель транзактов – вызовов.

б). OBMARSH – матрица (8*8) – общее число вызовов между парами узлов;

OTMARSH – матрица (8*8) – числа потерянных вызовов между парами узлов;

VMARSH – матрица (8*8) – доля потерянных вызовов по матрицам.

4. Лабораторное задание

4.1. Открыть программу моделирования МСС и сохранить её под именем 1МСС.

4.2. Запустить модель и зафиксировать основные результаты прогона, представленные в матрице Mitog.

4.3. Снять зависимости характеристик, представленных в матрице Mitog (см. п. 3.3) от следующих параметров:

- доля свободной полосы, выделяемой для передачи файлов (параметр doly);

- резерв полосы для приоритетных трафиков А и V (параметр rezp).

4.4. Провести оптимизацию системы по обобщённому критерию S (минимизация потерь) в зависимости одновременно от двух параметров, приведённых в п. 4.3 (doly и rezp).

4.5. Провести оптимизацию системы при резком изменении соотношения нагрузок по видам трафика. Например, трафиков речи и данных.

5. Содержание отчёта

Зафиксировать в табличном виде результаты прогонов по пунктам 4.3, 4.4 и 4.5 и построить графики этих зависимостей.

6. Контрольные вопросы

6.1. Опишите структуру моделируемой сети.

6.2. В чём особенности мультисервисных сетей.

6.3. В каком виде вводятся в модель параметры нагрузок по видам трафика (А, V и D).

6.4. Какая полоса пропускания выделяется в модели МСС для каждого вида трафика.

6.5. Объяснить ход кривых в графиках по пункту 4.3.

Список использованной литературы

1. Ю.Н. Корнышев, А.П. Пшеничников, А.Д. Харкевич. Теория телетрафика. – М.: Радио и Связь, 1996.
2. Гольдштейн Б.С., Соколов Н.А., Яновский Г.Г. Сети связи. Учебник для ВУЗов. БХВ-Петербург, 2010, 400 стр.
3. Л. Клейнрок. Вычислительные системы с очередями. – М.: Мир, 1979. – 600 с.
4. В.В. Крылов, С.С. Самохвалова. Теория телетрафика и ее приложения. – СПб.: ВHV-Санкт-Петербург, 2005.
5. Г.П. Башарин. Лекции по математической теории телетрафика. – М.: РУДН, 2009.
6. С.Н. Степанов. Основы телетрафика мультисервисных сетей. – М.: Эко-Трендз, 2010.
7. А.Н. Соколов, Н.А. Соколов. Однолинейные системы массового обслуживания. – Учебное пособие, СПбГУТ, 2010
8. В.М. Вишневский. Теоретические основы проектирования компьютерных сетей. Москва.: Техносфера, 2003.- 512с.
9. Мизин И.А., Богатырёв В.А., Кулешов А.П. Сети коммутации пакетов. – М.: Радио и связь, 1986. – 408 с.
- 10.Шварц М. Сети ЭВМ. Анализ и проектирование. – М.: Радио и связь, 1981. – 336 с.
- 11.Петров В.В. Структура телетрафика и алгоритм обеспечения качества обслуживания при влиянии эффекта самоподобия. Диссертация на соискание ученой степени к.т.н. М.: 2004.
- 12.Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. Учебник для ВУЗов. 3-е издание. Издательский дом - Питер. 2006. 958с.