

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
Некоммерческое акционерное общество
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ»

Кафедра Телекоммуникационных систем

«ДОПУЩЕН К ЗАЩИТЕ»

Зав.кафедрой
к.т.н., профессор
Байкенов А. С.

«___» _____ 2015 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ
пояснительная записка

на тему: Исследование процесса функционирования узла коммутации сети следующего поколения

Специальность 5В071900 Радиотехника, электроника и телекоммуникации

Выполнил магистрант гр. МТСП-13-1


(подпись) Талгаткызы Г.
(Ф.И.О.)

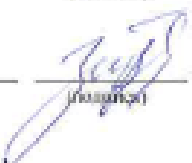
Научный руководитель к.т.н. профессор
(ученая степень, звание)


(подпись) Туманбаева К. Х.
(Ф.И.О.)

Консультант по ВТ к.т.н., старший преподаватель
(ученая степень, звание)


(подпись) Ефремова Ю.И.
(Ф.И.О.)

Нормоконтролер магистр ассистент
(ученая степень, звание)


(подпись) Зайцев Е.О.
(Ф.И.О.)

Рецензент к.т.н., доцент
(ученая степень, звание)

(подпись) Липская М. А.
(Ф.И.О.)

Алматы, 2015

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ»

Факультет «Радиотехники, электроники и связи»

Кафедра «Телекоммуникационных систем»

Специальность 6М071900 «Радиотехники, электроники и телекоммуникации»

ЗАДАНИЕ

на выполнение магистерской диссертации

Магистранту Талгаткызы Гаухар.

Тема диссертации Исследование процесса функционирования узла коммутации сети следующего поколения

Утверждена приказом по университету № 17 от «12» декабря 2014 г.

Срок сдачи законченной диссертации «_____» _____ 201__ г.

Перечень вопросов, подлежащих разработке в магистерской диссертации, или краткое содержание магистерской диссертации: предложена имитационная модель узла сети в случае поступления отдельных пакетов разных размеров и имитационная модель алгоритма управления очередями в узле сети с коммутацией пакетов с повторным поступлением в узел отброшенных пакетов

Перечень графического материала (с точным указанием обязательных чертежей):

1 Модуль RED

2 Зависимость вероятности отбрасывания пакетов механизмом RED от средней длины очереди;

3 Алгоритм управления очередями в узле сети с коммутацией пакетов

Основная рекомендуемая литература:

1 Боев В. Д. Моделирование систем. Инструментальные средства GPSS World. Санкт-Петербург: БХВ-Петербург, 2004.

2 Жерновой Ю.В. Простейшие модели управления очередью в узлах сетей с коммутацией каналов//Информационные процессы, том 14, №2, 2014

График
подготовки магистерской диссертации

Наименование разделов, перечень разрабатываемых вопросов	Сроки представления научному руководителю	Примечание
1 Качество обслуживания в сети следующего поколения	30.09.13 г.	
2 Архитектура сетевых механизмов обеспечения качества обслуживания качества обслуживания в сетях IP	15.10.13 г.	
3 Классификация алгоритмов семейства RED	07.01.14 .	
4 Модель узла сети в случае поступления отдельных пакетов разных размеров.	15.10.14 г.	
5 Анализ алгоритмов управления очередями в узле сети с коммутацией пакетов	10.11.14 г.	

Дата выдачи задания «__» _____ 201__ г.

Заведующий кафедрой _____ Байкенов А. С.
(подпись) (Ф.И.О.)

Руководитель диссертации _____ Туманбаева К. Х
(подпись) (Ф.И.О.)

Задание принял к исполнению магистрант _____ Талгаткызы Г.
(подпись) (Ф.И.О.)

Аңдатпа

Магистрлік диссертацияда келесі ұрпақ желісінің коммутация торабында кезекті басқару алгоритмдерін зерттеу жүргізілген. Пакеттердің желіден шығарылу мүмкіндіктерінің алгоритмдерін қолдану арқылы мәліметтер пакеттерін беру желілері тораптарының жұмысын моделдейтін қызмет жүйелері қарастырылған. GPSS World жүйесінде желіден шығарылған пакеттердің қайталанып түспейтін және олардың қайталанып түсуі алгоритмдері қолданылған мәліметтер пакетін беретін желілер торабының имитациялық моделдері қарастырылған.

Аннотация

В магистерской диссертации проведено исследование алгоритмов управления очередями в узле коммутации сети следующего поколения. Рассмотрены системы обслуживания, моделирующие работу узлов сетей передачи пакетов данных с использованием алгоритмов вероятностного отбрасывания пакетов. В системе GPSS World разработаны имитационные модели узла сети передачи пакетов данных с использованием алгоритма без повторного поступления в узел отброшенных пакетов и с их повторным поступлением.

Abstract

In master's dissertation investigated queuing algorithms in switching node next generation network. Modeling job of network nodes transmitting data packets using probabilistic algorithms dropping packets systems service is considered. In the system GPSS World were worked out the simulation models of node network of transmission of packages of data with the use of algorithm without the repeated entering knot of throw-away packages and with their repeated receipt

Содержание

Введение.....	6
1 Выбор предмета исследования.....	8
1.1 Качество обслуживания в сети следующего поколения.....	8
1.2 Работы МСЭ по стандартизации качества обслуживания в сетях IP	11
1.3 Архитектура сетевых механизмов обеспечения качества обслуживания качества обслуживания в сетях IP	15
1.4 Механизмы поддержки качество обслуживания в сетях IP	17
1.5 Модель предоставления интегрированных услуг (IntServ)	21
1.6 Модель предоставления дифференцированных услуг	22
1.7 Постановка задачи.....	24
2 Классификация алгоритмов семейства RED	25
2.1 Модель модуля RED	25
2.2 Построение классификации	28
2.3 Обзор алгоритмов на базе RED	29
2.3.1 Алгоритм Enhanced Random Early Detection	29
2.3.2 Адаптивные алгоритмы: ARED, RARED, POWARED	30
2.3.3 Алгоритм Double Slope RED	32
2.3.4 Алгоритм Gentle RED	32
2.3.5 Алгоритм State Dependent RED	33
2.3.6 Алгоритм Stabilized RED	34
2.3.7 Алгоритм Weighted RED	35
2.3.8 Семейство алгоритмов RED with In/Out	35
2.3.9 Алгоритм SF-RED	36
3 Модели управления очередью в узлах сетей с коммутацией пакетов.....	36
3.1 Одноканальная система с ограниченным буфером как модель поступления пакетов разных размеров	38
3.2 Стационарные характеристики одноканальной системы	39
3.3 Одноканальная система с ограниченным буфером	40
3.4 Многоканальная система без ограничений на длину очереди	42
3.5 Стационарные характеристики многоканальной системы	44
3.6 Случай ординарного входящего потока	47
3.7 Многоканальная система без ограничений на длину очереди	47
3.8 Анализ алгоритмов управления очередями в узле сети с коммутацией пакетов.....	49
Заключение	55
Список литературы	57
Приложение А	60
Приложение В.....	61
Приложение С.....	62

Введение

Одной из основных задач при построении сети следующего поколения является обеспечение требуемого уровня качества обслуживания различных видов трафика. С ростом объема информационных потоков в таких мультисервисных сетях возникают перегрузки в местах их концентрации, а именно в узлах коммутации сети. Под узлом коммутации будем подразумевать коммутатор, концентратор или маршрутизатор, то есть сетевое устройство с буфером, где пакеты ждут в очереди своего обслуживания. При перегрузке буфер переполняется, но увеличение его объема приводит к увеличению времени передачи пакета. То есть увеличение объема буфера снижает один из основных показателей качества обслуживания поступающего потока.

Для предотвращения перегрузок в узлах сетей с коммутацией пакетов используются алгоритмы активного управления очередью (Active Queue Management – AQM). Каждый пакет, поступающий в буфер узла коммутации сети, где он ждет обслуживания, может быть отброшен, даже если буфер не заполнен. Вероятность отбрасывания зависит от длины очереди, данную зависимость называют функцией отбрасывания. Задача анализа алгоритмов активного управления очередью с целью повышения качества обслуживания является актуальной.

В последние годы появились работы, в которых анализируется эффективность алгоритмов AQM с применением теории массового обслуживания. В качестве модели узла сети с коммутацией пакетов рассматривается многоканальная система массового обслуживания (СМО) с ожиданием и ограниченной длиной очереди. В данных работах не учитывается повторное поступление в узел отброшенных пакетов. Известно, что протоколы, работающие с установлением соединения, в случае если пакет отброшен и не доставлен получателю организуют повторную передачу пакета. Отправитель нумерует пакеты и ожидает от приемника служебное сообщение о том, что пакет получен. Время ожидания служебного извещения ограничено, если оно не получено за определенный промежуток времени, то пакет считается утерянным, и его отправляют повторно. Такой процесс называется квитированием, он является одним из методов обеспечения надежной связи.

Целью данной работы является анализ алгоритмов управления очередями в узле сети с коммутацией пакетов с учетом повторного поступления в узел отброшенных пакетов.

Для достижения поставленной цели поставлены следующие задачи:

- 1) анализ классификации алгоритмов активного управления очередями RED;
- 2) провести анализ моделей алгоритмов RED, применяющих теорию массового обслуживания;

3) разработка имитационную модель узла сети в случае поступления отдельных пакетов разных размеров;

4) разработка модель алгоритма управления очередями в узле сети с коммутацией пакетов без повторного поступления в узел отброшенных пакетов и модель алгоритма с повторным поступлением в узел отброшенных пакетов

Научная новизна работы заключается в том, что в ней исследованы модель алгоритма с повторным поступлением в узел отброшенных пакетов.

Практическая ценность работы заключается в использовании разработанных моделей для оценки эффективности алгоритмов управления очередями при их выборе для построения мультисервисных сетей.

1 Выбор предмета исследования

1.1 Качество обслуживания в сети следующего поколения

Качество обслуживания (Quality of Service, QoS) является предметом активных исследований и стандартизации на протяжении всей истории развития телекоммуникаций. Существенный вклад в развитие различных аспектов концепции QoS внес Международный союз электросвязи, включая, в том числе, разработку норм и требований к показателям качества обслуживания, стандартизацию сетевых механизмов, обеспечивающих необходимые показатели QoS, а также формулировку основополагающих определений.

Среди стандартов, посвященных качеству обслуживания в электросвязи, одно из центральных мест занимает рекомендация МСЭ E.800. В ней качество обслуживания определяется как «суммарный эффект рабочих характеристик обслуживания, который определяет степень удовлетворенности пользователя данной службой». Расширяя концепцию качества обслуживания, отвечающую Рекомендации E.800, Рекомендация МСЭ G.1000 разделяет рабочие характеристики обслуживания на функциональные компоненты и связывает их с сетевыми характеристиками, определенными в ряде рекомендации МСЭ – таких как I.350, Y.1540 и Y.1541.

В дополнение к Рекомендации МСЭ G.1000, определяющей структуру связей между рабочими характеристиками (производительностью, надежностью, потерями, задержкой и др.) и характеристиками сети, Рекомендация МСЭ G.1010 содержит спецификации требований со стороны приложений, ориентированных на конечного пользователя. Качество обслуживания нашло отражение в большом числе статей и книг, среди которых отметим монографии [1].

Исторически, первые системы оценок и механизмов поддержки качества обслуживания были разработаны для традиционных видов электросвязи – телеграфии и телефонии. Понятно, что сегодня при широком применении сетей передачи данных, быстром внедрении широкополосных технологий и замене телеграмм на сообщения электронной почты параметры качества обслуживания и механизмы их поддержки в телефонных сетях становятся все менее актуальными. При построении и эксплуатации ТфОП задача обеспечения гарантированного качества обслуживания состоит в том, чтобы обслуживание телефонного вызова осуществлялось с соблюдением всех установленных норм, в том числе, и заданных показателей качества передачи речи. Совокупность этих норм и соответствующих численных значений базируется на документах МСЭ и ETSI.

Модель услуг в ТфОП была основана на принципе установления соединения и в дальнейшем (70-е – 80-е годы прошлого столетия) была распространена на такие технологии передачи данных, как X.25, Frame Relay и широкополосные цифровые сети интегрального обслуживания (B-ISDN),

основанные на модели виртуальных каналов. В сетях В-ISDN рекомендации МСЭ (в частности, I.356 и I.610) и форума АТМ определяют параметры качества обслуживания и способы их измерений для межконцевых соединений.

В отличие от упомянутых выше технологии в классических сетях IP применяется метод доставки, полностью исключаящую любую форму организации соединений – как физических, так и виртуальных. Этот метод основан на рассылке пакетов-дейтаграмм. Качество доставки в традиционных сетях IP базируется на принципе так называемой «наилучшей попытки» (Best effort). Концепция «наилучшей попытки» предполагает, что пользователи справедливо разделяют доступные сетевые ресурсы, трафик передается со скоростью, максимально возможной в данных условиях загрузки ресурсов сети, но при этом не гарантируется обеспечение любого предварительно определенного уровня качества обслуживания. Очевидно, что такой подход к обслуживанию означает следующее: отсутствуют различия между разными видами трафика, нет гарантии в доставке пакетов в правильном порядке, и что он будет доставлен в требуемое время или вообще будет доставлен, и т.д.

Концепция «наилучшей попытки» была достаточно эффективной для приложений, где можно передавать данные не в реальном времени (электронная почта, передача файлов). Кроме того, с учетом переизбытка сетевых ресурсов в транспортных сетях, построенных на базе волоконно-оптических линий связи, принцип «наилучшей попытки» в определенной степени позволяет обеспечить сегодня требования телефонии (голос поверх IP) и других приложений реального времени [2].

Однако, как только возникает недостаток ресурсов, ведущий к увеличению вероятности потерь пакетов и росту их задержек, для приложений реального времени необходимые показатели качества обслуживания не могут быть обеспечены. Прежде всего, это объясняется основным принципом функционирования IP-сетей – передачей данных в дейтаграммном режиме, т.е. без установления соединений и без управления. С появлением новых приложений, особенно реального времени (интерактивная передача речи, видеотелефония и видеоконференции), вопрос о гарантированном качестве обслуживания в сетях IP становится одним из наиболее сложных. Это объясняет, почему качество обслуживания в сетях IP считается предметом постоянного внимания МСЭ, ETSI, IETF и других организаций стандартизации и электросвязи.

Сегодня общепризнано, что сети с коммутацией каналов и пакетов постепенно эволюционируют в направлении создания общей инфраструктуры, базирующейся на протоколах семейства IP. Этот процесс получил название конвергенции. Инфраструктура, возникшая в результате конвергенции, должна будет обеспечивать транспортировку трафика телефонных сетей, сетей телевидения и трафика приложений, традиционно использующих сети Интернет. Подобный сценарий конвергенции предлагает как экономический

выигрыш, получаемый благодаря объединению технологий, так и определяет развитие сектора телекоммуникаций через создание новых услуг.

Однако, процесс конвергенции до настоящего времени протекает достаточно медленно. И здесь мы вновь возвращаемся к проблеме обеспечения необходимого качества обслуживания, которая является одним из основных тормозящих факторов в процессе конвергенции сетей и услуг и построения единой сети на базе IP, рассматриваемой сегодня как сеть следующего поколения (Next Generation Network, NGN). Чтобы полностью реализовать преимущества конвергенции в будущих IP-ориентированных сетях, необходимо разработать новые принципы распределения ресурсов сетей и управления трафиком, которые будут гарантировать различные уровни показателей качества обслуживания для большого и разнообразного числа приложений, реализуемых конечными пользователями.

При этом разделение ресурсов и процессы управления трафиком должны быть скоординированы в условиях наличия большого числа разнообразных приложений с существенно отличающимися требованиями к рабочим характеристикам сети (таблица 1.1). Детальное рассмотрение рабочих характеристик, определяющих качество обслуживания, и соответствующих норм будет проведено в следующих разделах.

Таблица 1.1 – Чувствительность различных приложений к сетевым характеристикам

Тип трафика	Уровень чувствительности к сетевым характеристикам			
	Полоса пропускания	Потери	Задержка	Джиттер
Голос	Очень низкий	Средний	Высокий	Высокий
Электронная коммерция	Низкий	Высокий	Высокий	Низкий
Транзакции	Низкий	Высокий	Высокий	Низкий
Электронная почта	Низкий	Высокий	Низкий	Низкий
Telnet	Низкий	Высокий	Средний	Низкий
Поиск в сети «от случая к случаю»	Низкий	Средний	Средний	Низкий
Постоянный поиск в сети	Средний	Высокий	Высокий	Низкий
Пересылка файлов	Высокий	Средний	Низкий	Низкий
Видеоконференция	Высокий	Средний	Высокий	Высокий
Мультикастинг	Высокий	Высокий	Высокий	Высокий

1.2 Работы МСЭ по стандартизации качества обслуживания в сетях IP

В рамках работ МСЭ по стандартизации качества обслуживания в сетях IP предполагаются следующие этапы решения задачи обеспечения QoS для сетей, построенных на базе IP-ориентированных протоколов:

- создание согласованного общего набора рабочих характеристик сетей IP и норм для него;
- внедрение сетевых механизмов, которые будут обеспечивать заданные показатели качества обслуживания и конфигурации «терминал-терминал»;
- вложение нормированных значений показателей качества обслуживания в протоколы сигнализации;
- разработка архитектуры сетевых механизмов поддержки.

В 2002 г. ИК 13 МСЭ-Т опубликовала два международных стандарта, которые отвечают первому из перечисленных этапов. Рекомендация МСЭ Y.1540 описывает стандартные сетевые характеристики для передачи пакетов в сетях IP. Рекомендация МСЭ Y.1541 [5] определяет нормы для параметров, определенных в Y.1540, между двумя граничными сетевыми интерфейсами – точками подключения оконечных терминальных устройств. Кроме того, в этой рекомендации специфицированы шесть классов качества обслуживания в зависимости от приложений [3].

Эти рекомендации важны для всех участников телекоммуникационного сценария-операторов и провайдеров, производителей оборудования и конечных пользователей. Сетевые операторы и провайдеры будут использовать их при планировании, развертывании и оценке сетей IP в соответствии с требованиями конечных пользователей к качеству обслуживания. Производители будут опираться на эти рекомендации при создании оборудования, которое должно отвечать спецификациям сетевых провайдеров. Наконец, конечные пользователи (в первую очередь, корпоративные) смогут применить рекомендации Y.1540 и Y.1541 при оценке характеристик реально функционирующих IP-сетей с позиций соответствия этих характеристик требованиям потребителей. Рассмотрим некоторые детали рекомендаций Y.1540 и Y.1541, касающиеся основных сетевых характеристик, связанных с обеспечением QoS в сетях IP.

Рекомендация МСЭ Y.1540

В рекомендации Y.1540 рассматриваются следующие сетевые характеристики, как наиболее важные по степени их влияния на сквозное качество обслуживания (от источника до получателя), оцениваемое пользователем:

- производительность сети;
- надежность сети/сетевых элементов;
- задержка;

- вариация задержки (джиттер);
- потери пакетов.

Производительность сети (или скорость передачи данных) пользователя определяется как эффективная скорость передачи, измеряемая в битах в секунду. Следует отметить, что значение этого параметра не совпадает с максимальной пропускной способностью сети, ошибочно называемой (причем, довольно часто) полосой пропускания. Минимальное значение производительности обычно гарантируется провайдером услуг, который, в свою очередь должен иметь соответствующие гарантии от сетевого провайдера.

В рекомендации Y.1540 не приведены нормативные характеристики производительности сети, которые различаются для различных приложений. Вместе с тем, в Рекомендации Y.1541 отмечено, что параметры, связанные с эффективной скоростью передачи, могут быть определены через дескриптор трафика IP-сети, описанный в рекомендации МСЭ Y.1221.

Надежность сети/сетевых элементов. Пользователи обычно ожидают высокий уровень надежности от систем связи. Надежность сети может быть определена через ряд параметров, из которых наиболее часто используется коэффициент готовности, вычисляемый как отношение времени простоя объекта к суммарному времени наблюдения объекта, включающему время простоя и время между отказами. В идеальном случае коэффициент готовности должен быть равен 1, что означает стопроцентную готовность сети. На практике коэффициент готовности составляет 0,999, что соответствует 9 часам времени недоступности (простоя) сети в год. Готовность сети ТфОП оценивается величиной «пять девяток», что означает 5,5 мин. простоя в год. В таблице 1.2 приведены данные по времени простоя для различного количества «девяток».

Таблица 1.2 – Коэффициент готовности и соответствующие значения времени простоя оборудования

Коэффициент готовности	Время простоя
0,99	3,7 дней в год
0,999	9 часов в год
0,9999	53 минуты в год
0,99999	5,5 минуты в год
0,99999999	30 секунд в год

Необходимо отметить, что обеспечение коэффициента готовности «пять девяток» в сетях IP, построенных на традиционном оборудовании данных (серверы, маршрутизаторы), являются достаточно серьезной проблемой. Причина этого состоит в том, что обработка информационных потоков в сетях IP в значительной части базируется на программном обеспечении (а не аппаратном, как это имеет место в ТфОП). В то же время статистика отказов

сетевого оборудования показывает, что надежность программного обеспечения примерно в два раза ниже надежности аппаратного обеспечения.

Параметры доставки пакетов IP. В общем случае сеанс связи состоит из трех фаз – установления соединения, передачи информации и разъединения соединения. В рекомендации Y.1540 из трех фаз сеанса рассматривается только вторая – фаза доставки пакетов IP. Такой подход отражает природу сетей IP, не ориентированных на установление соединений. Спецификацию рабочих характеристик и параметров QoS для двух других фаз (установление и разъединение соединения) планируется провести в дальнейшем.

Рекомендация МСЭ-Т Y.1540 определяет следующие параметры, характеризующие доставку IP-пакетов [4].

Задержка доставки пакета IP (IP packet transfer delay, IPTD). Параметр IPTD определяется как время $(t_2 - t_1)$ между двумя событиями – вводом пакета во входную точку сети в момент t_1 , где $(t_2 > t_1)$ и $(t_2 - t_1) \leq T_{\max}$.

В общем, параметр IPTD определяется как время доставки пакета между источником и получателем для всех пакетов – как успешно переданных, так и пораженных ошибками.

Средняя задержка доставки пакета IP - параметр, специфицированный в Рекомендации Y.1540, определяется как средняя арифметическая величина задержек пакетов в выбранном наборе переданных и принятых пакетов. Значение средней задержки зависит от передаваемого в сети трафика и доступных сетевых ресурсов, в частности, от пропускной способности. Рост нагрузки и уменьшение доступных сетевых ресурсов ведут к росту очередей в узлах сети и, как следствие к увеличению средних задержек доставки пакетов.

Речевая информация и, отчасти, видеoinформация являются примерами трафика, чувствительного к задержкам, тогда как приложения данных в основном менее чувствительны к задержкам. Когда задержка доставки пакета превышает определенные значения T_{\max} , такие пакеты отбрасываются. В приложениях реального времени (например, в IP-телефонии) это ведет к ухудшению качества речи. Ограничения, связанные со средней задержкой пакетов IP, играют ключевую роль для успешного внедрения технологии Voice over IP (VoIP), видео-конференций и других приложений реального времени. Этот параметр во многом будет определять готовность пользователей принять подобные приложения.

Вариация задержки пакета IP (IP packet delay variation, IPDV). Параметр V_k , характеризует вариацию задержки IPDV. Для IP-пакета с индексом k этот параметр определяется между входной и выходной точками сети в виде разности между абсолютной величиной задержки X_k при доставке пакета с индексом k , и определенной эталонной (или опорной) величиной задержки доставки пакета IP, $d_{1,2}$, для тех же сетевых точек:

$$V_k = X_k - d_{1,2} \quad (1.1)$$

Эталонная задержка доставки пакета IP, $d_{1,2}$ между источником и

получателем определяется как абсолютное значение задержки доставки первого пакета IP между данными сетевыми точками. Вариация задержки пакета IP, или джиттер, проявляется в том, что последовательные пакеты прибывают к получателю в нерегулярные моменты времени. В системах IP-телефонии это, к примеру, ведет к искажениям звука и в результате к тому, что речь становится неразборчивой.

И определенной этолонной (или опорной) величиной задержки доставки пакета

Коэффициент потери пакетов IP (IP packet loss ratio, IPLR). Коэффициент IPLR определяется как отношение суммарного числа потерянных пакетов к общему числу принятых в выбранном наборе переданных и принятых пакетов. Потери пакетов в сетях IP возникают в том случае, когда значение задержек при их передаче превышает нормированное значение, определенное выше T_{max} . Если пакеты теряются, то при передаче данных возможна их повторная передача по запросу принимающей стороны. В системах VoIP пакеты, пришедшие к получателю с задержкой, превышающей T_{max} , отбрасываются, что ведет к провалам в принимаемой речи. Среди причин, вызывающих потери пакетов, необходимо отметить рост очередей в узлах сети, возникающих при перегрузках.

Коэффициент ошибок пакетов IP (IP packet error ratio, IPER). Коэффициент IPER определяется как суммарное число пакетов, принятых с ошибками, к сумме успешно принятых и пакетов, принятых с ошибками.

Рекомендация МСЭ Y.1541

Рекомендация Y.1540 определяет численные значения параметров, специфицированных в ней, которые должны выполняться в сетях IP на международных трактах, соединяющих терминалы пользователей. Нормы на параметры разделены по различным классам QoS, которые определены в зависимости от приложений и сетевых механизмов, применяемых для обеспечения гарантированного качества обслуживания. В таблице 1.3 [5] представлены нормы на определенные выше сетевые характеристики.

Значения параметров, приведенные в таблице, представляют собой, соответственно, верхние границы для средних задержек, джиттера, потерь и ошибок пакетов. В Рекомендации Y.1541 представлены спецификации набора параметров, которые связаны с измерением реальных значений сетевых характеристик – периода наблюдений, длины тестовых пакетов, числа пакетов и т.д. Рекомендуемый интервал измерений для задержки, джиттера и потерь составляет не менее 60 с.

Рекомендация Y.1541 устанавливает соответствие между классами качества обслуживания и приложениями:

- Класс 0 – приложения реального времени, чувствительные к джиттеру, характеризующиеся высоким уровнем интерактивности (VoIP, видеоконференции).
- Класс 1 - приложения реального времени, чувствительные к джиттеру, интерактивные (VoIP, видеоконференции).

- Класс 2 – транзакции данных характеризуемые высоким уровнем интерактивности (например, сигнализация).
- Класс 3 – транзакции данных, интерактивные.
- Класс 4 – приложения, допускающие низкий уровень потерь (короткие транзакции, массивы данных, потоковое видео).
- Класс 5 – традиционные применения сетей IP.

Таблица 1.3 – Нормы для характеристик сетей IP с распределением по классам качества

Сетевые характеристики	Классы QoS					
	0	1	2	3	4	5
Задержка доставки пакета IP, IPTD	100 мс	400 мс	100 мс	400 мс	1с	Н
Вариация задержки пакета IP, IPDV	50мс	50мс	Н	Н	Н	Н
Коэффициент потери пакетов IP, IPLR	1×10^{-3}	1×10^{-3}	1×10^{-3}	1×10^{-3}	1×10^{-3}	Н
Коэффициент ошибок пакетов IP, IPER	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}	Н

Примечание: Н- не нормировано

1.3 Архитектура сетевых механизмов обеспечения качества обслуживания качества обслуживания в сетях IP

Помимо определения сетевых параметров и спецификации норм для них ИК 13 МСЭ-Т проводит в настоящее время работы по идентификации и стандартизации сетевых механизмов, обеспечивающих QoS в IP-ориентированных сетях. В мае 2004 г. была принята Рекомендация МСЭ Y .1291 [6]. Описывающая архитектурную модель для поддержки качества обслуживания в сетях с пакетной передачей.

Сетевые механизмы должны использоваться в комбинации с характеристиками качества обслуживания, формируемыми в зависимости от приложений. При разработке архитектуры сетевых механизмов учитывалось, что различные услуги будут иметь разнообразные требования к характеристикам сети. Например, для телемедицины точность доставки играет более существенную роль, чем суммарная средняя задержка или джиттер,

тогда как для IP-телефонии джиттер и задержка являются ключевыми характеристиками и должны быть минимизированы.

С учетом тенденции постоянного расширения числа приложений с различными требованиями к характеристикам качества обслуживания архитектура поддержки QoS должна включать в себя широкий набор общих сетевых механизмов, как существующих, так и перспективных, подлежащих разработке.

Архитектура поддержки QoS определяет набор сетевых механизмов, называемых конструктивными блоками. В настоящее время определен начальный набор конструктивных блоков, отвечающих трем логическим плоскостям: плоскости контроля, плоскости данных (информационной плоскости) и плоскости административного управления (см.рисунок).

Плоскость контроля. Механизмы QoS контрольной плоскости оперируют с путями, по которым передается трафик пользователей, и включают в свой состав:

- управление допуском (Admission Control, AC)
- маршрутизацию для QoS (QoS routing)
- резервирование ресурсов (Resource reservation)

Плоскость данных. Эта группа механизмов оперирует непосредственно с пользовательским трафиком и включает себя:

- управление буферами (Buffer management)
- предотвращение перегрузок (Congestion avoidance)
- маркировку пакетов (Packet marking)
- организацию и диспетчеризацию очередей (Queuing and scheduling)
- формирование трафика (Traffic shaping)
- правила обработки трафика (Traffic policing)
- классификацию трафика (Traffic classification)

Плоскость административного управления. Эта плоскость содержит механизмы QoS, имеющие отношение к эксплуатации, администрированию и управлению сетью применительно к доставке пользовательского трафика. В число механизмов QoS на этой плоскости входят:

- измерения (Metering)
- заданные правила доставки (Policy)
- восстановление трафика (Traffic restoration)
- соглашение об уровне обслуживания (Service Level Agreement)

Сетевые механизмы QoS (или, следуя терминологии МСЭ, блоки QoS) могут быть специфицированы применительно к сетевым узлам (например, управление буферами узлов) или к сетевым сегментам (маршрутизация QoS), где понятие “сетевой сегмент” может относиться к межконцевому соединению, участку доступа, межузловому участку или участку, соединяющему две и более сетевой. Далее мы рассмотрим некоторые из перечисленных выше механизмов.

Для иллюстрации того, как различные механизмы поддержки QoS могут быть использованы в стандартизованных методах обеспечения требуемых показателей качества обслуживания, мы рассмотрим два наиболее широко применяемых в настоящее время подхода при решении задачи обеспечения качества обслуживания: интегрированные (IntServ) и дифференцированные услуги (DiffServ).

1.4 Механизмы поддержки качества обслуживания в сетях IP

Как было отмечено выше, переход к сетям следующего поколения, построенным на базе стека протоколов IP, возможен только при условии, что для большого числа приложений будут обеспечены соответствующие показатели качества обслуживания. Для достижения этой цели был разработан ряд механизмов борьбы с задержками и потерями, которые в соответствии с разрабатываемой Рекомендацией МСЭ-ТУ.1291 разделены по трем плоскостям-плоскости контроля, плоскости данных и плоскости административного управления (рисунок 1.1).

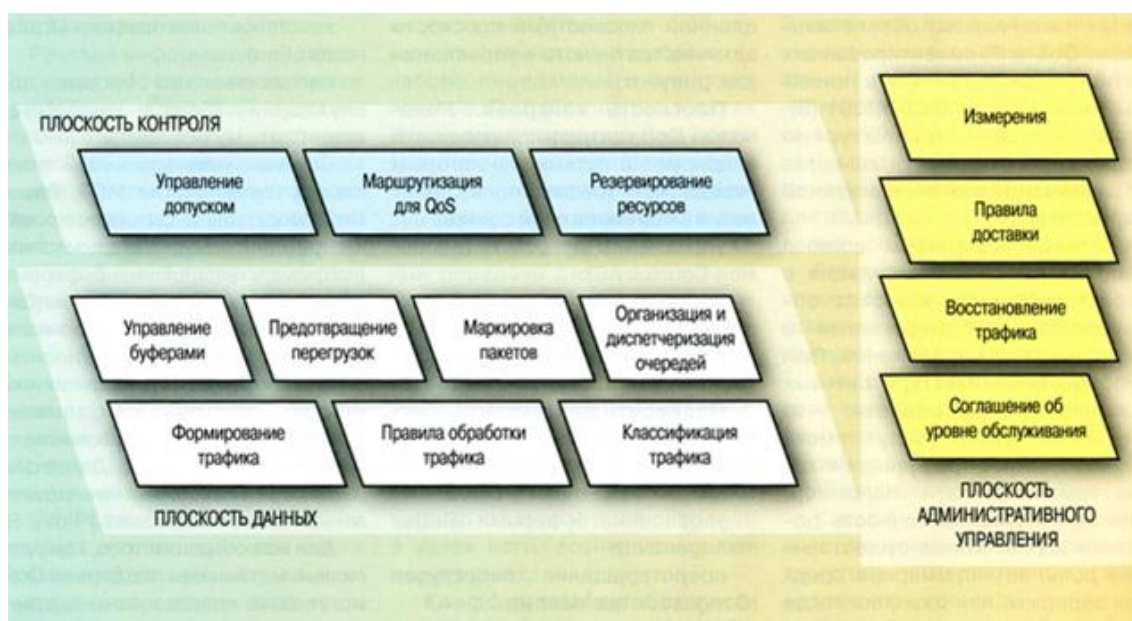


Рисунок 1.1 – Плоскости контроля, данных, административного управления

Управления допуском (Call Admission Control). Этот механизм контролирует новые заявки на пропуск трафика через сеть, определяя, может ли вновь поступающий трафик привести к перегрузке сети или к ухудшению уровня качества обслуживания для уже имеющегося в сети трафика. Обычно управление допуском построено на определенном наборе правил администрирования, контроля и управления сетевыми ресурсами.

Эти правила могут быть специфицированы в соответствии с потребностями сетевого провайдера или базироваться на соглашении между

провайдером и пользователем и включать в свой состав различные параметры QoS. Для удовлетворения требований определенных служб (например, при чрезвычайных обстоятельствах), соответствующему трафику может быть присвоен высший приоритет при доступе в сеть.

Маршрутизация QoS (QoS routing) обеспечивает выбор пути, который удовлетворяет требованиям к качеству обслуживания для конкретного потока данных. Выбираемый путь может отличаться от кратчайшего пути. Процесс определения пути предполагает знание требований к качеству обслуживания со стороны потока данных и наличие информации о доступных сетевых ресурсах.

В настоящее время предложено большое число возможных методов определения наилучшего пути по критерию QoS. Как правило, в вычислениях наилучшего пути маршрутизации QoS учитывается либо одна сетевая характеристика, либо две (производительность и задержка, стоимость и производительность, стоимость и задержка и т.д.), с тем, чтобы сделать процесс вычислений приемлемым для инженерных расчетов.

Резервирование ресурсов (Resource reservation). В целом, необходимым условием для обеспечения резервирования ресурсов является наличие ресурсов в сети. Резервирование ресурсов широко использовалось в сетях АТМ при формировании постоянных виртуальных соединений. В IP-ориентированных сетях наиболее типичным механизмом резервирования является механизм, базирующийся на протоколе RSVP.

Механизмы QoS в плоскости данных

Управление буферами (Buffer management). Управление буферами (или очередями) состоит в управлении пакетами, стоящими в узлах в очереди на передачу. Основные задачи управления очередями – минимизация средней длины очереди при одновременном обеспечении высокого использования канала, а также справедливое распределение буферного пространства между различными потоками данных. Схемы управления очередями различаются, в основном, критерием, по которому отбрасываются пакеты, и местом в очереди, откуда производится сбор пакетов (начало или конец очереди). Наиболее простым критерием для сброса пакетов является достижение очередью определенного порога, называемого максимальной длиной очереди.

Более распространены сегодня так называемые механизмы активного управления очередями. Типичным примером является алгоритм RED (Random Early Detection – раннее случайное обнаружение перегрузки). При использовании алгоритма RED поступающие в буфер пакеты сбрасываются на основании оценки средней длины очереди.

Предотвращение перегрузок (Congestion avoidance). Механизмы предотвращения перегрузок поддерживают уровень нагрузки в сети ниже ее пропускной способности. Обычный способ предотвращения перегрузок состоит в уменьшении трафика, поступающего в сеть. Как правило, команда уменьшить из примеров механизм предотвращения перегрузок является механизм окна в протоколе TCP.

Маркировка пакетов (Packet marking). Пакеты могут быть промаркированы в соответствии с определенным классом обслуживания. Маркировка обычно производится во входном пограничном узле, где специальное поле заголовка (Type of Service в заголовке IP или DS-байт в заголовке DiffServ, см. ниже) вводится определенное значение. Кроме того, маркировка применяется для тех пакетов, которые могут быть удалены в случае перегрузки сети.

Организация и планирование очередей (Queuing and scheduling). Цель механизмов этой группы – выбор пакетов для передачи из буфера в канал. Большинство дисциплин обслуживания (или планировщиков) основаны на схеме “первый пришел – первый обслуживается”. Для обеспечения более гибких процедур ввода пакетов из очереди был предложен ряд схем, основанных на формировании нескольких очередей. Среди них, в первую очередь, необходимо назвать схемы приоритетного обслуживания. Другой пример гибкой организации очереди – механизм взвешенной справедливой буферизации (Weighted Fair Queuing, WFQ), когда ограниченная пропускная способность на выходе узла распределяется между несколькими потоками (очередями) в зависимости от требования к пропускной способности со стороны каждого потока.

Еще одна схема организации очереди основана на классификации потоков по классу обслуживания (Class-Based Queuing, CBQ). Потоки классифицируют в соответствии с классами обслуживания и затем размещают в буфере в различных очередях. Каждой очереди выделяется определенный процент выходной пропускной способности в зависимости от класса, и очереди обслуживаются по циклической схеме.

Формирование трафика (Traffic shaping). Формирование или управление характеристиками трафика предполагает контроль скорости передачи пакетов и объема потоков, поступающих на вход сети. В результате прохождения через специальные формирующие буферы уменьшается пачечность исходного трафика, и его характеристики становятся более предсказуемыми. Известны два механизма обработки трафика – Leaky Bucket (“дырявое дерево”) и Token Bucket (“ведро с жетонами”). Алгоритм Leaky Bucket регулирует скорость пакетов, покидающих узел. Независимо от скорости входного потока, скорость на выходе узла является величиной постоянной.

В противоположность этому, алгоритм Token Bucket не регулирует скорость на выходе узла и не сбрасывает пакеты. Скорость пакетов на выходе узла может быть такой же, как и на входе, если только в соответствующем накопителе (“ведре”) есть жетоны. Жетоны генерируются с определенной скоростью и накапливаются в ведре. Алгоритм характеризуется двумя параметрами – скоростью генерации жетонов и размером памяти (размером “ведра”) для них. Пакеты не могут покинуть узел, если в ведре нет жетонов. И наоборот, сразу пачка пакетов может покинуть узел, израсходовав соответствующее число жетонов.

Правила обработки трафика (Traffic classification). Классификация трафика может быть проведена на потоковом или пакетном уровне. На входе в сеть в узле доступа (пограничном маршрутизаторе) пакеты классифицируются для того, чтобы выделить пакеты одного потока, характеризуемого общими требованиями к качеству обслуживания. Затем трафик подвергается процедуре нормирования (механизм Traffic Conditioning). Нормирование трафика предполагает измерение его параметров и сравнение результатов с параметрами, оговоренным в контракте по трафику, известному как соглашение об уровне обслуживания (Service Level Agreement, SLA, см. ниже). Если условия SLA нарушаются, то часть пакетов может быть отброшена. Магистральные маршрутизаторы, составляющие ядро сети, обеспечивают пересылку пакетов в соответствии с требуемым уровнем QoS.

Механизмы QoS в плоскости административного управления

Измерение (Metering). Измерение обеспечивает контроль параметров трафика – например, скорость потока данных в сравнении с согласной в SLA скоростью. По результатам измерений могут быть реализованы определенные процедуры – такие, как сброс пакетов и применение механизмов Leaky Bucket и Token Bucket.

Заданные правила доставки (Policy). Под правилами доставки здесь понимается набор правил, используемых для контроля и административного управления доступом к сетевым ресурсам. На основе таких правил поставщики услуг могут осуществлять реализацию механизмов в плоскости данных. Возможными применениями правил доставки являются маршрутизация по заданным правилам, фильтрация пакетов на основе заданных правил (маркировка или отбрасывание пакетов), регистрация заданных потоков, правила обработки, связанные с безопасностью.

Восстановление трафика (Traffic restoration). Под восстановлением трафика в данной рекомендации понимается реакция сети, смягчающая последствия в условиях отказа. Восстановление трафика рассматривается на различных уровнях эталонной модели процессов. На физическом уровне при использовании SDH надежность обеспечивается автоматической защитной коммутацией. На канальном уровне транспортных сетей восстановление трафика обеспечивается специальными механизмами, развитыми для кольцевых и ячеистых структур. Соответствующие процедуры предусмотрены в технологии ATM. Восстановление на сетевом уровне (протокол IP) осуществляется с помощью технологии MPLS.

Соглашение об уровне обслуживания (Service Level Agreement). Одним из основных понятий в концепции обеспечения требуемого уровня качества обслуживания в современных сетях является соглашение об уровне обслуживания. Первые SLA-контракты были разработаны в середине 90-х годов при предоставлении услуг передачи данных с использованием технологий Frame Relay, ATM и IP. Необходимости подобных контрактов была вызвана возрастающим требованием к оператором со стороны клиентов, чей бизнес все больше зависел от надежной и своевременной передачи

информации. Контракт SLA предполагает повышенную ответственности поставщика услуг, дисциплинирует его. В какой-то степени это дисциплинирует и заказчика, поскольку заключению соглашения предшествует этап анализа требований к уровню сервиса.

Соглашение SLA, называемое в ряде источников контрактом по трафику, представляет собой контракт между пользователем и провайдером услуг/сетевым провайдером. В контракте определяются основные характеристики (профиль) трафика, формируемого в оборудовании пользователя, и параметр QoS, предоставляемые провайдером. Соглашение SLA может включать в себя также и ценовые характеристики. Техническая часть SLA специфицирует набор параметров и их значения, которые вместе определяют уровень обслуживания, обеспечиваемый трафику пользователя со стороны сетевого провайдера.

Контракт SLA может быть статическим (согласовывается на длительный период – месяц, год и т.п.) или динамическим (определяется для каждого сеанса). В последнем случае для запроса требуемого уровня QoS должен использоваться сигнальный протокол (например, RSVP). Соглашения SLA, прежде всего, предполагают четко регламентированные обязательства поставщика услуг по обеспечению их качества (время предоставления услуг, например, круглосуточно или только в рабочие дни; время реакции на инцидент; время выезда персонала к заказчику; время закрытия инцидента и т.д.), а так же штрафные санкции за нарушение регламента. Из опыта зарубежных сетевых провайдеров известно, что стоимость SLA добавляется к стоимости гарантийного обслуживания и в ряде случаев может быть в несколько раз выше стоимости гарантийного обслуживания.

1.5 Модель предоставления интегрированных услуг (IntServ)

Процесс превращения сети Интернет в середине 90-х из академической в коммерческую инфраструктуру, рост числа узлов и количества пользователей, применение для разнообразных приложений с различными требованиями к качеству обслуживания – все эти факторы определили быстрое развитие механизмов поддержки QoS. В ответ на новые условия, возникшие в сетях IP, Комитет IETF предложил большой набор моделей и механизмов для обеспечения качества обслуживания в сетях Интернет, которые разделяются на две категории в соответствии с названиями рабочих групп.

Комитет IETF, разрабатывающих эти модели и механизмы – интегрированных услуг и дифференцированных услуг.

Рабочая группа Integrated Services Working Group разрабатывали модель предоставления интегрированных услуг (или IntServ), основанную на принципе интегрированного резервирования ресурсов. Модель IntServ была разработана для поддержки приложений реального времени, чувствительных к задержкам. Механизмы, реализующие модель интегрированных услуг,

должны обеспечивать взаимодействие всех сетевых устройств для поддержки любого уровня QoS вдоль пути передачи определенного потока пакетов.

Наиболее детально среди механизмов группы IntServ проработан протокол RSVP (Resource ReSerVation Protocol), спецификация которого (RFC 2205, [7]) была принята Комитетом IETF в 1997 г. Механизмы группы IntServ относятся к группе методов, гарантирующих “жесткое” или абсолютное качество обслуживания. Протокол RSVP является наиболее известным представителем группы механизмов интегрированного обслуживания. По существу, RSVP представляет собой протокол сигнализации, в соответствии с которым осуществляется резервирование и управление ресурсами с целью гарантии “жесткого” качества обслуживания.

Резервирование производится для определенного потока IP-пакетов перед началом передачи этого потока. Идентификация потока (определение пакетов, принадлежащих одному потоку) осуществляется по специальной метке, размещаемой в основном заголовке каждого пакета IPv6. После резервирования пути начинается передача пакетов данного потока, обслуживаемых на всем межконцевом соединении с заданным качеством.

Протокол RSVP является только протоколом сигнализации. Для обеспечения требуемого качества обслуживания на фазе переноса пакетов трафика он должен быть дополнен одним из существующих протоколов маршрутизации, а также набором механизмов управления трафиком, включающих управление допустимостью соединений (CAC), классификацию трафика, управление и планирование очередей, а также другие механизмы, составляющие основу архитектуры механизмов поддержки QoS, рассмотренную выше.

Несмотря на возможности протоколов группы IntServ в плане обеспечения требуемых показателей QoS, реализация и развертывание методов интегрированного обслуживания связаны с определенными трудностями, особенно в территориально распределенных сетях. В частности, необходимо учитывать возможность перегрузки маршрутизаторов и переполнения накопителей в сетевых узлах при большом числе одновременно обслуживаемых потоков. Необходимо также признать, что протоколы группы IntServ не отвечают требованиям масштабируемости.

1.6 Модель предоставления дифференцированных услуг

Модель дифференцированных услуг (Differentiated Services, Diff-Serv) является логическим продолжением работ IETF над архитектурой IntServ. Недостатки, заложенные в самом принципе модели IntServ (жесткие гарантии качества обслуживания, низкий уровень масштабирования) привели к необходимости создания более гибких механизмов обеспечения QoS. Общая характеристика принципов предоставления дифференцированных услуг [RFC-2475, [8]) была опубликована в декабре 1998г., а более детальные спецификации появились в середине 1999 г. Методы Diff-Serv составляют

группу механизмов, которые в отличие от методов IntServ обеспечивают относительное или “мягкое” качество обслуживания.

Основная идея механизмов Diff-Serv состоит в предоставлении дифференцированных услуг для набора классов трафика, отличающихся требованиями к показателям качества обслуживания. Как и в случае механизмов IntServ, для реализации дифференцированных услуг широко применяются механизмы, входящие в состав рассмотренной выше архитектуры поддержки QoS в сетях IP.

Одним из центральных понятий модели Diff-Serv является соглашение об уровне обслуживания, входящее в состав механизмов QoS на плоскости менеджмента. В модели Diff-Serv архитектура сети представляется в виде двух сегментов - пограничных участков и ядра. На входе в сеть в узле доступа (пограничном маршрутизаторе) пакеты классифицируются (механизм Traffic classification) для того, чтобы выделить пакеты одного потока, характеризуемого общими требованиями к качеству обслуживания. Затем трафик подвергается процедуре нормирования (механизм Traffic conditioning). Нормирование трафика предполагает измерение его параметров и сравнение результатов с параметрами, оговоренным в контракте SLA. Если условия SLA нарушаются, то часть пакетов может быть отброшена. При необходимости поток пакетов проходит через устройство профилирования (механизм Traffic shaping). Магистральные маршрутизаторы, составляющие ядро сети, обеспечивают пересылку пакетов в соответствии с требуемым уровнем QoS.

Требования к необходимому набору показателей качества обслуживания задаются в специальном однобайтовом поле каждого пакета – в октете Type of Service (ToS) протокола IPv4 или в октете Traffic Class (TC) протокола IPv6. Отметим, что в модели Diff-Serv это поле называется DS-байтом.

Содержание DS-байта определяет вид предоставляемых услуг. Первые два бита определяют приоритет пакета, следующие четыре – требуемый класс обслуживания пакета в узле, и два бита остаются неиспользуемыми. Класс обслуживания здесь означает механизм обработки и продвижения пакета из данного узла к следующему (Per-Hop Behavior, PHB) в соответствии с необходимым качеством обслуживания. Таким образом, с помощью поля DS можно определить до 32 различных уровней качества обслуживания.

В стандартах IETF RFC 2598 и RFC 2597 были определены два класса услуг для модели Diff-Serv. В спецификации RFC 2598 описан класс “срочной доставки” (Expedited Forwarding, EF), обеспечивающий наивысший из возможных уровней качества обслуживания (Premium Service) и применяемый для приложений, требующих доставки с минимальными значениями задержки и джиттера.

Второй класс обслуживания, получивший название “гарантированной доставки” (Assured Forwarding, AF), представлен в спецификации RFC 2597. Класс гарантированной доставки поддерживает уровень качества обслуживания более низкий, чем класс срочной доставки, но более высокий,

чем обслуживание по принципу “наилучшей попытки” (Best effort). Внутри этого диапазона QoS класс AF определяет четыре типа трафика и три уровня отбрасывания пакетов. Таким образом, класс AF обеспечивает возможность обслуживания до 12 разновидностей трафика в зависимости от набора требуемых показателей качества обслуживания.

Обработка пакетов в соответствии с определенными уровнем приоритета и типом трафика осуществляется специальными схемами обслуживания очередей, обеспечивающими контроль задержек и джиттера пакетов и исключение возможных потерь.

Среди основных механизмов управления очередями отметим приоритетное обслуживание (Priority Queuing), взвешенное справедливое обслуживание (Weighted Fair Queuing) и обслуживание в соответствии с механизмом PNB (Class-Based Queuing).

Относительная простота классификации трафика в модели Diff-Serv и отсутствие механизмов сквозного (end-to-end) резервирования ресурсов определяют широкие возможности применения дифференцированных услуг по сравнению с механизмами IntServ. Применение механизмов Diff-Serv в магистральном ядре сети позволяет использовать их для обработки агрегированного трафика, который может объединяться в пограничных сегментах сети. Такой подход может оказаться эффективным, например, в IP-телефонии, когда множество речевых потоков объединяются в один агрегированный, характеризуемый одинаковыми требованиями к показателям качества обслуживания.

По-видимому, механизмы Diff-Serv все же не могут гарантировать такой же уровень QoS, какой можно получить в цифровых телефонных сетях, базирующихся на коммутации каналов (например, в ISDN). Вместе с тем, можно ожидать, что в будущих сетях доля служб, требующих такой уровень качества, будет относительно небольшой, тогда как для приложений с менее критическими требованиями к QoS модели и механизмы дифференцированных услуг будут способны обеспечить необходимый уровень качества обслуживания.

1.7 Постановка задачи

Целью данной работы является анализ алгоритмов управления очередями в узле сети с коммутацией пакетов с учетом повторного поступления в узел отброшенных пакетов.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) провести анализ классификации алгоритмов активного управления очередями RED;
- 2) провести анализ моделей алгоритмов RED, применяющих теорию массового обслуживания;

- 3) разработать имитационную модель узла сети в случае поступления отдельных пакетов разных размеров;
- 4) разработать модель алгоритма управления очередями в узле сети с коммутацией пакетов без повторного поступления в узел отброшенных пакетов и модель алгоритма с повторным поступлением в узел отброшенных пакетов

2 Классификация алгоритмов семейства RED

2.1 Модель модуля RED

Алгоритмы регулирования состояния потока обычно реализованы в виде модулей в сетевом оборудовании (например, маршрутизаторе). Функционирование модуля, реализующего алгоритм типа RED, можно схематично представить следующим образом (рисунок 2.1).

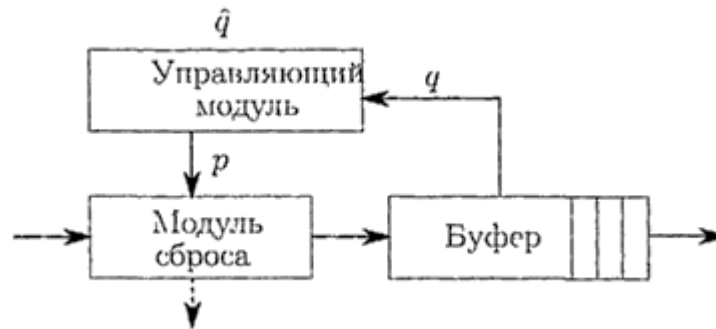


Рисунок 2.1- Модуль RED

Пакет передаваемых данных при поступлении в систему попадает в модуль сброса. Решение о сбросе пакета принимается на основе значения функции $\rho(\hat{q})$, получаемого от управляющего модуля. Функции $\rho(\hat{q})$ зависит от значения экспоненциально взвешенного скользящего среднего размера длины очереди \hat{q} , также вычисляемого управляющим модулем, основываясь на текущем значении длины очереди q .

Классический алгоритм RED подробно рассмотрен в работе [9]. Здесь приведены лишь формулы для расчета функции сброса $\rho(\hat{q})$ и экспоненциально взвешенной скользящей средней длины очереди \hat{q} .

Для вычисления \hat{q} используется рекуррентная формула экспоненциально взвешенного скользящего среднего (Exponentially Weighted Moving-Average, EWMA):

$$\hat{q}_{k+1} = (1 - \omega_q)\hat{q}_k + \omega_q q_k, \quad k = 0, 1, 2, \dots, \quad (2.1)$$

где $\omega_q, 0 < \omega_q < 1$ – весовой коэффициент экспоненциально взвешенного скользящего среднего. В работе [10] предложена следующая формула для определения ω_q :

$$\omega_q = 1 - e^{-1/c}, \quad (2.2)$$

где C – пропускная способность канала (пакетов в секунду).

Функция $\rho(\hat{q})$ сброса пакетов линейно зависит от \hat{q} , минимального q_{min} и максимального q_{max} пороговых значений и параметра максимального сброса ρ_{max} , задающего максимальный уровень сброса пакетов при достижении \hat{q} значения q_{max} , и вычисляется следующим образом:

$$\rho(\hat{q}) = \begin{cases} 0, & 0 < \hat{q} \leq q_{min}, \\ \frac{\hat{q} - q_{min}}{q_{max} - q_{min}} \rho_{max}, & q_{min} < \hat{q} \leq q_{max}, \\ 1, & \hat{q} > q_{max}. \end{cases} \quad (2.3)$$

Под сбросом в данном случае понимается как собственно отбрасывание пакета в соответствии с функцией $\rho_d(\hat{q})$, так и маркировка пакета в соответствии с функцией $\rho_m(\hat{q})$ для последующей отсылки сообщения ECN (Explicit Congestion Notification) [11]. Таким образом, функция сброса представима в виде $\rho(\hat{q}) = \rho_d(\hat{q}) + \rho_m(\hat{q})$.

Применение функции сброса к поступившему в систему пакету может носить случайный характер. В этом случае вероятность ρ_a того, что за время X между моментами сброса пакетов будет сброшено n пакетов, может иметь геометрическое распределение с параметром $\rho(\hat{q})$:

$$\rho_a = P\{X = n\} = (1 - \rho(\hat{q}))^{n-1} \rho(\hat{q}) \quad (2.4)$$

или равномерное распределение:

$$\rho_a = \begin{cases} \frac{\rho(\hat{q})}{1 - (n-1)\rho(\hat{q})} \prod_{i=0}^{n-2} \left(1 - \frac{\rho(\hat{q})}{1 - i\rho(\hat{q})}\right) = \rho(\hat{q}), & 1 \leq n \leq \frac{1}{\rho(\hat{q})} \\ 0, & n > \frac{1}{\rho(\hat{q})}. \end{cases} \quad (2.5)$$

Для второго случая вероятность быть сброшенным для каждого входящего пакета равна $\rho(\hat{q}) / (1 - count \cdot \rho(\hat{q}))$, где $count$ – количество несброшенных пакетов, прибывших после последнего сброшенного.

Далее рассмотрим, каким изменениям можно подвергнуть классический алгоритм RED. Поведение модуля RED (рисунок 2.1) можно описать системой

$$\begin{cases} \hat{q} = \hat{Q}(\rho), \\ \rho(\hat{q}) = P(\hat{q}), \end{cases} \quad (2.6)$$

образованной функцией экспоненциально взвешенной скользящей средней длины очереди $\hat{q} = \hat{Q}(\rho)$ и функцией сброса $\rho(\hat{q}) = P(\hat{q})$. Решение (\hat{q}_s, ρ_s) системы уравнений (2.6) задает ее равновесное состояние (рисунок 2.2). Пороговые значения q_{min} и ρ_{max} задают рабочую область системы (2.6). Алгоритм эффективно действует только в ее границах. При этом выход за верхнюю границу менее желателен, чем за нижнюю.

Для эффективной работы алгоритма типа RED его функция сброса $\rho(\hat{q}) = P(\hat{q})$ должна удовлетворять следующим условиям:

- в районе нижней границы рабочего диапазона ее наклон (т.е. первая производная) должен быть как можно меньше (для повышения эффективности использования полосы пропускания);
- в районе верхней границы рабочего диапазона ее наклон должен быть максимальным (для предотвращения выхода системы из рабочего диапазона).

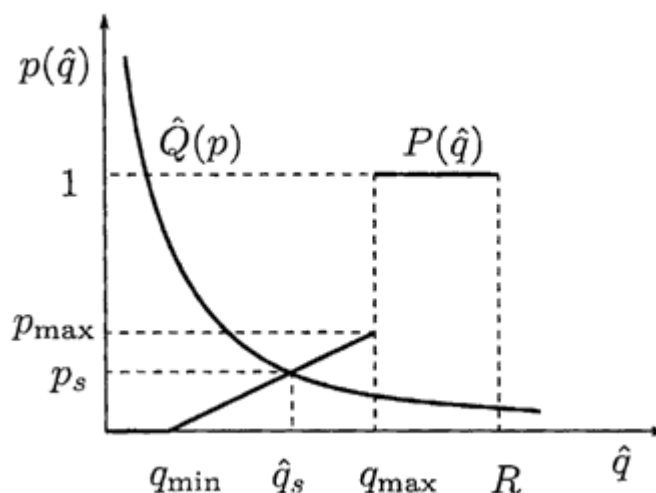


Рисунок 2.2 - Точка равновесия для системы (1.2)

Из этого можно сделать вывод, что функция сброса должна быть нелинейной. Но основным требованием при разработке алгоритма RED стало обеспечение малых накладных расходов при обработке трафика, что проявилось в простоте всех операций в этом алгоритме. Он построен так, чтобы минимизировать сложные расчеты (умножение, вычисления с плавающей точкой). Поэтому при модификации базового алгоритма вместо нелинейной функции сброса можно использовать ее аппроксимацию кусочно-линейной функцией [12].

Функция экспоненциально взвешенной скользящей средней длины очереди $\hat{q} = \hat{Q}(\rho)$ зависит от многих параметров (характеристик потока, дисциплины обслуживания и т.д.) и требует построения математической модели для своего описания. Можно модифицировать формулу (2.1), считая, что вся зависимость от внешних параметров учитывается мгновенной длиной

очереди q . При этом \hat{q} должна строиться так, чтобы противодействовать выходу системы из рабочей области.

2.2 Построение классификации

При построении классификации использован фасетный метод, подразумевающий параллельное разделение множество объектов на независимые классификационные группы. Классы обозначены прописными латинскими буквами, подклассы – в скобках после символа класса. Характеристики классификации:

- количество модулей RED – класс M;
- алгоритмы с единым для всех модулей RED буфером – подкласс CS (Complete Sharing);
- алгоритмы со своим (виртуальным) буфером у каждого модуля – подкласс CP (Complete Partitioning);
- вид функции $\hat{Q}(\rho) - Q$;
- вид функции сброса $P(\hat{q})$ - класс P.
- нелинейная в рабочей области функция – подкласс NL (Non-Linear);
- нелинейная с кусочно-линейной аппроксимацией в рабочей области функция – подкласс NLA (Non-Linear with Approximation);
- измененная вне рабочей области функция, возвращающая систему в устойчивое состояние – подкласс RS (Return to Steady);
- с дополнительной аддитивной или мультипликативной функцией сброса – подклассы DA (Additive Drop) и DM (Multiplicative Drop) соответственно.

Предложенную структуру классификации можно пояснить следующим образом.

Модуль RED может обрабатывать лишь один поток трафика. Разновидности с несколькими блоками RED применимы в случае, когда имеется несколько потоков, которые должны быть обработаны с разными параметрами. Обычно такая ситуация возникает в системах DiffServ, предоставляющих дифференцированное обслуживание классам трафика с разным требованиям к качеству обслуживания (подход к построению и анализу моделей управления трафиком в IP- сетях с предоставлением дифференцированных услуг предложен в соавторстве автором диссертации в работах[13-14]). При этом параметры RED (пороговые значения и параметр максимального сброса) могут быть едиными для всех типов трафика или индивидуальными для каждого потока.

Как уже было сказано, поведение модуля RED описывается функциями $P(\hat{q})$ и $\hat{Q}(\rho)$. Изменения экспоненциально взвешенной скользящей средней длины очереди (2.1) направлены на то, чтобы увеличить крутизну в областях значений \hat{q} , больших верхнего порога. Это обусловлено тем, что при выходе за пределы рабочей области система должна скорее вернуться в рабочий диапазон.

Наибольший интерес представляют алгоритмы, изменяющие вид функции сброса. Предположение о предпочтительности нелинейного вида функции сброса уже обсуждалось выше. Также было сказано о необходимости уменьшения вычислительной сложности алгоритма, что можно достигнуть аппроксимацией нелинейной функции сброса кусочно-линейной функцией.

Изменение вида функции сброса за пределами рабочей области служит, как и в случае с функцией $\hat{Q}(\rho)$, возвращению системы в устойчивое состояние.

При необходимости задать дополнительную зависимость функции сброса от других параметров (например, явно учитывать количество входящих потоков) можно задать дополнительную функцию сброса, которая может быть аддитивной: $\rho(\hat{q}) = P_{RED}(\hat{q}) + P_{add}$, или мультипликативной: $\rho(\hat{q}) = P_{RED}(\hat{q}) \times P_{add}$.

Построенная таким образом классификация априори является частичной, поскольку примененный фасетный метод не задает полной таксономии предметной области. Данная классификация ортогональна в том смысле, что конкретный алгоритм можно отнести одновременно к нескольким пунктам классификации. Для иллюстрации построенной классификации далее приведено описание наиболее характерных для каждого алгоритмов семейства RED.

2.3 Обзор алгоритмов на базе RED

2.3.1 Алгоритм Enhanced Random Early Detection

В алгоритме Enhanced Random Early Detection (ERED), предложенном в работе [15], изменениям подвергается функция экспоненциально взвешенной скользящей средней длины очереди (рисунок 2.3):

$$\hat{q}_{k+1} = \begin{cases} \frac{1-w_q}{\alpha} \hat{q}_k + w_q q_k, & \hat{q}_k \leq q_{min}, \alpha > 1, \\ (1-w_q) \hat{q}_k + w_q q_k, & q_{min} \leq \hat{q} < q_{max}, k = 0,1,2, \dots, \\ (1-w_q) \hat{q}_k + \frac{w_q q_k}{\beta}, & \hat{q}_k \geq q_{max}, \beta > 1, \end{cases} \quad (2.7)$$

где α и β – некоторые коэффициенты, позволяющие управлять задержкой реагирования \hat{q} на изменение текущей длины очереди q , что уменьшает амплитуду и частоту колебаний экспоненциально взвешенной скользящей средней длины очереди \hat{q} .

В этом алгоритме поведение функции $\hat{Q}(\rho)$ модифицируется таким образом, чтобы при выходе за пределы рабочей области значение \hat{q} как можно скорее в нее возвратилось. Алгоритм соответствует классу Q.

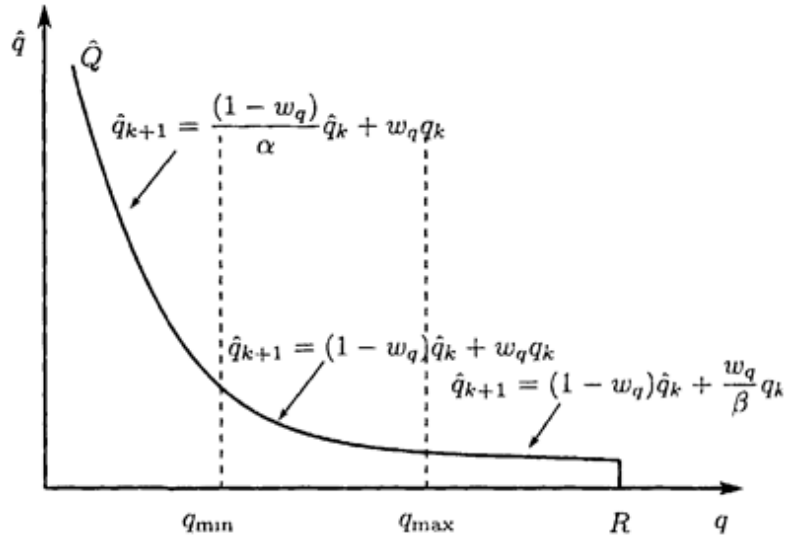


Рисунок 2.3 – Функция экспоненциально взвешенной скользящей средней длины очереди в алгоритме ERED

2.3.2 Адаптивные алгоритмы: ARED, RARED, POWARED

Как было отмечено ранее, сложно подобрать оптимальный набор параметров RED. Адаптивные механизмы позволяют динамически изменять параметры алгоритма в зависимости от характера поведения трафика.

В алгоритме Adaptive RED (ARED), разработанном Фенгом [16], функция сброса $\rho(\hat{q})$ модифицируется посредством изменения по принципу MIMD (Multiplicative Increase Multiplicative Decrease) параметра ρ_{max} таким образом, чтобы удерживать \hat{q} между заданными минимальным q_{min} и максимальным q_{max} порогами (т.е. $\rho_{max} = \rho_{max}(\hat{q})$).

Данный алгоритм был усовершенствован С. Флорид и теперь под алгоритмом ARED подразумевают именно эту модификацию. В ней используется принцип AIMD (Additive Increase Multiplicative Decrease) [17].

Алгоритм ARED функционирует следующим образом. Для каждого интервала времени, выступающего в качестве параметра алгоритма, если значение \hat{q} больше целевого (желаемого) \hat{q}_t значения и $\rho_{max} \leq 0,5$, то ρ_{max} увеличивается на величину α ; в противном случае, если $\hat{q} < \hat{q}_t$ и $\rho_{max} \geq 0,01$, то ρ_{max} уменьшается в β раз:

$$\rho_{max} = \begin{cases} \rho_{max} + \alpha, & \hat{q} > \hat{q}_t, \\ \rho_{max} \cdot \beta, & \hat{q} < \hat{q}_t, \end{cases} \quad \begin{matrix} \rho_{max} \leq 0,5, \\ \rho_{max} \geq 0,01, \end{matrix} \quad (2.8)$$

где

$$\alpha = \min\left(0,01; \frac{\rho_{max}}{4}\right), \quad \beta = 0,9$$

$$q_{min} + 0,4(q_{max} - q_{min}) < \hat{q}_t < q_{min} + 0,6(q_{max} - q_{min}).$$

В выбор верхней границы 0,5 для ρ_{max} обуславливается тем, что при $\rho_{max} = 0,5$ значение $\rho(\hat{q})$ меняется от 0 до 1 при изменении \hat{q} с q_{min} до $2q_{max}$, таким образом, обеспечивается достаточно хорошая производительность и в том случае, когда интенсивность сброса пакетов больше 50%.

Нижняя граница ρ_{max} со значением 0,01 объясняется в [18] желанием ограничить изменения ρ_{max} : при $q_{max} = kq_{min}$ (k – коэффициент, показывающий, во сколько раз верхний порог больше нижнего) $\hat{q}_t = \frac{k+1}{2} q_{min}$. При интенсивности сброса пакетов, приближающейся к 100%, и $\rho_{max} = 50\%$ значение \hat{q} приближается к величине $2q_{max}$. На увеличение ρ_{max} от 0,01 до 0,5 требуется $0,49/\alpha$ интервалов (максимальное увеличение ρ_{max} за один интервал времени может произойти на величину $\min(0,01; \frac{\rho_{max}}{4})$). Аналогично для уменьшения ρ_{max} с 0,5 до 0,01 требуется $\frac{\log 0,02}{\log \beta}$ интервалов.

Таким образом, посредством изменения ρ_{max} строится нелинейная функция сброса $\rho(\hat{q})$. Алгоритм относится к классу P(NL).

Алгоритм Refined ARED (RARED) [19] является модификацией ARED и предлагает более активно изменять ρ_{max} для более быстрой адаптации $\rho(\hat{q})$ к изменению \hat{q} :

$$\rho_{max} = \begin{cases} \rho_{max} + \alpha, & \hat{q} > \hat{q}_t, \\ \beta \rho_{max}, & \hat{q} \leq \hat{q}_t, \end{cases} \quad \begin{matrix} \rho_{max} \leq 0,5, \\ \rho_{max} \geq 0,01, \end{matrix} \quad (2.9)$$

где

$$q_{min} + 0,48(q_{max} - q_{min}) < \hat{q}_t < q_{min} + 0,52(q_{max} - q_{min}),$$

$$\alpha = \left(0,25 \frac{\hat{q} - \hat{q}_t}{\hat{q}_t}\right) \rho_{max}, \quad \beta = 1 - \left(0,17 \frac{\hat{q}_t - \hat{q}}{\hat{q}_t - q_{min}}\right)$$

Алгоритм POWARED [20] также является модификацией алгоритма ARED. Его суть заключается в том, что интервал $Q = [q_{min}; q_{max}]$ ($\hat{q} \in Q$) разбивается на 3 части: $Q_1 = [q_{min}; \hat{q}_t)$, $Q_2 = (\hat{q}_t; q_{max}]$ и $Q_3 = \hat{q}_t$, а ρ_{max} вычисляется по формуле:

$$\rho_{max} = \begin{cases} \rho_{max} - \delta_1, & \hat{q} \in Q_1, \\ \rho_{max} + \delta_2 & \hat{q} \in Q_2, \\ \rho_{max}, & \hat{q} \in Q_3. \end{cases} \quad (2.10)$$

где $\delta_1 = \left| \left(\frac{\hat{q} - \hat{q}_t}{\beta \hat{q}_t} \right)^k \right|$, $\delta_2 = \left| \left(\frac{\hat{q}_t - \hat{q}}{\beta (q_{max} - \hat{q}_t)} \right)^k \right|$, $\hat{q}_t = \frac{1}{2}(q_{max} + q_{min})$, $K(K > 1, K = 2, 3, \dots)$ – множитель, β – параметр сжатия. Здесь используется принцип AIAD (Additive Increase Additive Decrease).

2.3.3 Алгоритм Double Slope RED

В алгоритме Double Slope Random Early Detection (DSRED) введено дополнительное пороговое значение q_{mid} между q_{min} и q_{max} , причем $q_{mid} = 0,5(q_{max} + q_{min})$ [21]. Функция сброса $\rho(\hat{q})$ описывается двумя линейными сегментами с углами наклона α и β соответственно, регулируемые задаваемым селектором режимов γ (рисунок. 2.4):

$$\rho(\hat{q}) = \begin{cases} 0, & \hat{q} < q_{min}, \\ \alpha(\hat{q} - q_{min}), & q_{min} \leq \hat{q} < q_{mid}, \\ 1 - \gamma + \beta(\hat{q} - q_{mid}), & q_{mid} \leq \hat{q} < q_{max}, \\ 1, & \hat{q} \geq 2q_{max}, \end{cases} \quad (2.11)$$

где $\alpha = \frac{2(1-\gamma)}{q_{max} - q_{min}}$, $\beta = \frac{2\gamma}{q_{max} - q_{min}}$, а \hat{q} вычисляется так же, как и в алгоритме RED по формуле (2.1).

Таким образом, идеей данного алгоритма является задание нелинейной функции сброса, аппроксимированной ломаной линией с двумя звеньями. Алгоритм относится к классу P(NLA).

2.3.4 Алгоритм Gentle RED

Функция сброса $\rho(\hat{q})$ в алгоритме Gentle RED (GRED) [22] имеет вид (рисунок. 2.5):

$$\rho(\hat{q}) = \begin{cases} 0, & \hat{q} < q_{min}, \\ \frac{\hat{q} - q_{min}}{q_{max} - q_{min}} \rho_{max}, & q_{min} \leq \hat{q} < q_{max}, \\ \frac{\hat{q} - q_{max}}{q_{max}} (1 - \rho_{max}) + \rho_{max}, & q_{max} \leq \hat{q} < 2q_{max}, \\ 1, & \hat{q} \geq 2q_{max}, \end{cases} \quad (2.12)$$

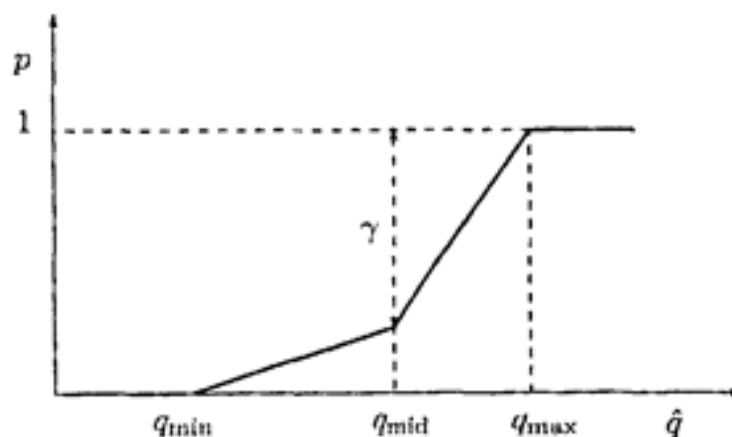


Рисунок 2.4 – Вид функции сброса в алгоритме DSRED

При формальной схожести поведения функции сброса алгоритмов DSRED и GRED между ними существует принципиальное различие. В DSRED функция имеет вид ломаной линии в пределах рабочего диапазона (q_{min}, q_{max}), фактически аппроксимируя нелинейную функцию.

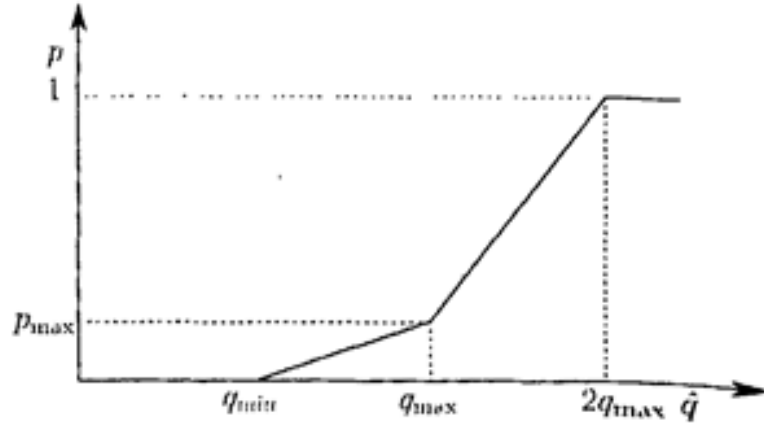


Рисунок 2.5 – Вид функции сброса в алгоритме GRED

В GRED в рабочем диапазоне находится стандартная функция сброса RED. Дополнительный же сегмент служит для обеспечения более плавного поведения функции сброса за пределами рабочего диапазона. Алгоритм относится к классу P(RS).

2.3.5 Алгоритм State Dependent RED

В алгоритм SDRED изменяются максимальный порог q_{max} и вес w_q очереди. Функция сброса $\rho(\hat{q})$ (рисунок 2.6) имеет вид [23]:

$$\rho(\hat{q}) = \begin{cases} 0, & \hat{q} < q_{min}, \\ \frac{\hat{q} - q_{min}}{q_{max} - q_{min}} \rho_{max}, & q_{min} \leq \hat{q} < q_{max}, \\ \frac{\hat{q} - q_{min}}{(q_{max} + 0,1R) - q_{min}} \rho_{max}, & q_{max} \leq \hat{q} < 0,7R, \\ \frac{\hat{q} - q_{min}}{(q_{max} + 0,2R) - q_{min}} \rho_{max}, & 0,7R \leq \hat{q} < 0,8R, \\ \frac{\hat{q} - q_{min}}{(q_{max} + 0,3R) - q_{min}} \rho_{max}, & 0,8R \leq \hat{q} < 0,9R, \\ 1, & 0,9R \leq \hat{q} \leq R, \end{cases} \quad (2.13)$$

$$\hat{q}_{i+1} = \begin{cases} (1 - w_q) \hat{q}_i + w_q q_i, & \hat{q}_i < q_{max}, 0,9R \leq \hat{q}_i \leq R, \\ (1 - k w_q) \hat{q}_i + k w_q q_i, & q_{max} \leq \hat{q}_i < 0,7R, \\ (1 - k^2 w_q) \hat{q}_i + k^2 w_q q_i, & 0,7R \leq \hat{q}_i < 0,8R, \\ (1 - k^3 w_q) \hat{q}_i + k^3 w_q q_i, & 0,8R \leq \hat{q}_i < 0,9R, \end{cases} \quad (2.14)$$

где k – константа для изменения w_q в зависимости от коэффициента занятости очереди.

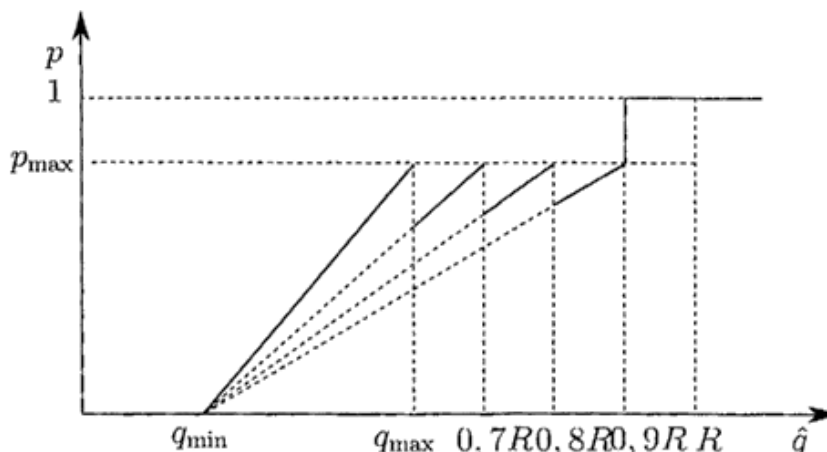


Рисунок 2.6 – Вид функции сброса в алгоритме SDRED

Экспоненциально взвешенная скользящая средняя длина очереди вычисляется следующим образом:

$$\hat{q}_{i+1} = \begin{cases} (1-w_q)\hat{q}_i + w_q q_i, & \hat{q}_i < q_{max}, 0,9R \leq \hat{q}_i \leq R, \\ (1-kw_q)\hat{q}_i + kw_q q_i, & q_{max} \leq \hat{q}_i < 0,7R, \\ (1-k^2w_q)\hat{q}_i + k^2w_q q_i, & 0,7R \leq \hat{q}_i < 0,8R, \\ (1-k^3w_q)\hat{q}_i + k^3w_q q_i, & 0,8R \leq \hat{q}_i < 0,9R, \end{cases} \quad (2.14)$$

где k – константа для изменения w_q в зависимости от коэффициента занятости очереди.

В данном алгоритме применяются сразу два подхода $\hat{Q}(\rho)$ и функции сброса $P(\hat{q})$. Оба эти подхода направлены на ускорение возврата \hat{q} в рабочую область. Алгоритм относится к классу QP(RS).

2.3.6. Алгоритм Stabilized RED

Алгоритм SRED [24] использует дополнительную функцию сброса, зависящую от числа активных соединений или потоков. С помощью данного алгоритма выявляются потоки, требующие неоправданно много ресурсов по сравнению с другими потоками. Эта схема позволяет держать занятость буфера на заранее определенных уровнях в независимости от количества активных соединений.

Результирующая функция представляет собой произведение полученной дополнительной функции сброса RED, то есть алгоритм относится к классу P(DM).

Отметим, что для ступенчатой функции сброса $p(\hat{q})$ и $w_q = 1$, \hat{q} становится эквивалентной q .

2.3.7. Алгоритм Weighted RED

Взвешенный алгоритм случайного раннего обнаружения (Weighted Random Early Detection, WRED) позволяет отдельно задать параметры (пороговые значения, максимальный уровень сброса) потокам разного типа трафика, обеспечивая таким образом различные уровни обслуживания пакетов в зависимости от уровня их отбрасывания. Алгоритм WRED работает с единой очередью пакетов, для которой, как и в RED, по формуле (2.1) рассчитывается экспоненциально взвешенное скользящее среднее. Пороговые значения длины очереди для различных уровней сброса пакетов могут быть полностью не совпадающими, частично перекрывающимися или полностью совпадающими (рисунок 2.7). Алгоритм относится к классу M (CS).

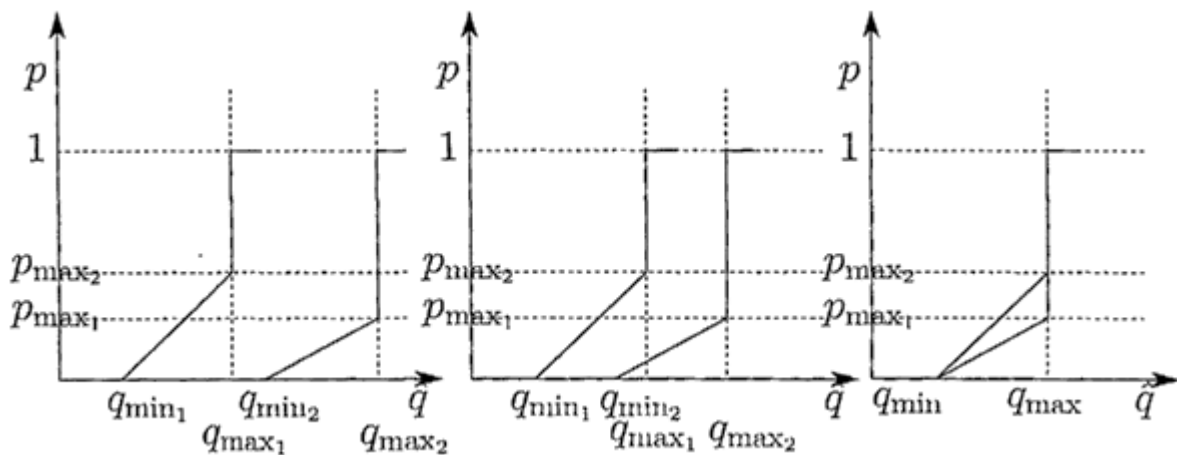


Рисунок 2.7- Вид функции сброса в алгоритме WRED

2.3.8 Семейство алгоритмов RED with In / Out

В алгоритме RED with In/Out (RIO, случайное раннее обнаружение с профильными/непрофильными пакетами) [25] поступающие пакеты делятся на профильные (IN) и непрофильные (OUT). Пакеты поступающего трафика определяются как IN-пакеты, если трафик находится в пределах некоторой заданной политики, и как OUT-пакеты, если трафик вышел за её пределы.

Для принятия решения о сбросе OUT-пакетов используется алгоритм RED относительно экспоненциально взвешенной скользящей средней длины общей очереди, а для принятия решения о сбросе IN-пакетов – алгоритм RED относительно экспоненциально взвешенной скользящей средней длины виртуальной очереди только из IN-пакетов.

В алгоритме RED with IN/Out Coupled Virtual Queues (RIO – C₁ случайное раннее обнаружение с профильными/непрофильными пакетами и

парными виртуальными очередями) используется различные ограничения в зависимости от приоритета обслуживания трафика. Принятие решения о сбросе IN – пакетов с приоритетом $j, j < n$ зависит от значения экспоненциально взвешенной скользящей средней длины виртуальной очереди, состоящей только из IN-пакетов с приоритетом, меньшим или равным j . Принятие решения о сбросе OUT – пакетов с приоритетом n зависит от средней занятости общей (физической) очереди.

В алгоритме RED with In/Out and De – Coupled Queues (RIO – DC, случайное раннее обнаружение с профильными/непрофильными пакетами и непарными виртуальными очередями) значение экспоненциально взвешенной скользящей средней длины очереди для пакетов определённого типа рассчитывается на основе их числа в очереди.

2.3.9. Алгоритм SF-RED

Алгоритм SF-RED [26] использует одну физическую и i (в соответствии с числом классов трафика) виртуальных очередей, для каждой из которых заданы свои пороговые значения q_{min_i} и q_{max_i} , а функция сброса вычисляется по формуле

$$\rho(\hat{q}) = \begin{cases} 0, & \hat{q}_i < q_{min_i}, \\ \frac{\hat{q}_i - q_{min_i}}{q_{max_i} - q_{min_i}}, & q_{min_i} \leq \hat{q}_i < q_{max_i}, \\ 1, & \hat{q}_i > q_{max_i}. \end{cases} \quad (2.15)$$

где \hat{q}_i – экспоненциально взвешенная скользящая средняя длина виртуальной очереди i , зависящая от изменяемого значения веса w_{q_i} .

Для каждого интервала времени и класса i

$$q_{min_i} = \begin{cases} q_{min_i} - \alpha, & \hat{q}_i > \hat{q}_t, \\ q_{min_i} + \alpha, & \hat{q}_i < \hat{q}_t, \end{cases} \quad (2.16)$$

$$\rho_{max_i} = \begin{cases} \rho_{max_i} + \beta, & \frac{\hat{q}_i}{\hat{q}_1} < w_{q_{ir}}, i \neq 1 \\ \rho_{max_i} - \beta, & \frac{\hat{q}_i}{\hat{q}_1} > w_{q_{ir}}, i \neq 1 \end{cases} \quad (2.17)$$

где \hat{q}_t – целевое значение экспоненциально взвешенной скользящей средней длины физической очереди, $\alpha = 1, \beta = 0,1, q_t = [(1 + 10\%) \hat{q}_t, (1 - 10\%) \hat{q}_t], w_{q_{ir}} = [(1 + 10\%) w_{q_{ir}}, (1 - 10\%) w_{q_{ir}}]$.

Алгоритм относится к классу M(CP) P(NL,DM).

В сводной таблице 2.1 приведены распределения рассмотренных алгоритмов семейства RED по классам.

Таблица 2.1 – Распределение алгоритмов семейства RED по классам

Класс	M		Q	P				
	CS	CP		NL	NLA	RS	DA	DM
WRED	+	-	-	-	-	-	-	-
RIO	-	+	-	-	-	-	-	-
SF-RED	-	+	-	+	-	-	-	+
ARED	-	-	-	+	-	-	-	-
RARED	-	-	-	+	-	-	-	-
POWARED	-	-	-	+	-	-	-	-
ERED	-	-	+	-	-	-	-	-
DSRED	-	-	-	-	+	-	-	-
GRED	-	-	-	-	-	+	-	-
SDRED	-	-	+	-	-	+	-	-
SRED	-	-	-	-	-	-	-	+

3 Модели управления очередью в узлах сетей с коммутацией пакетов

С целью предотвращения перегрузок в узлах сетей с коммутацией пакетов (ATM, TCP/IP и др.) используются алгоритмы активного управления очередью (англ. Active Queue Management – AQM). В соответствующей системе обслуживания, моделирующей работу узла сети, каждый поступающий пакет может быть отброшен с определенной вероятностью, даже если буфер еще полностью не заполнен. Вероятность отбрасывания зависит от длины очереди в момент поступления пакета. Зависимость вероятности отбрасывания пакетов от длины очереди называют функцией отбрасывания [27].

В алгоритмах AQM используют различные функции отбрасывания, например в известном алгоритме RED (Random Early Detection – случайное раннее обнаружение) [28]. Эта функция является линейной. Благодаря профилактическому случайному отбрасыванию пакетов алгоритм AQM косвенно информирует отправителя, использующего протокол TCP, о приближающейся перегрузке. Применение такого алгоритма в маршрутизаторе может принести много полезных эффектов, в том числе сокращение очереди и времени сетевой задержки (больше подробностей можно найти в [29]).

Результаты исследований показывают [27], что механизм функции отбрасывания является мощным средством для регулирования параметров системы обслуживания. С его помощью можно регулировать не только длину очереди, вероятность потери заявок, время ожидания, дисперсию длины очереди, но и несколько из этих параметров одновременно. Модели с функцией

отбрасывания имеют также свой глубокий универсальный смысл. Во всех случаях, когда нет возможности регулировать параметры системы обслуживания через изменения входящего потока либо процесса обслуживания, применение функции отбрасывания является простым и эффективным способом для обеспечения требуемых параметров системы обслуживания.

Использование теории массового обслуживания для анализа алгоритмов активного управления очередью началось в последние годы [30]. Как правило, авторы ограничивались изучением одноканальных систем с ограниченным буфером.

Системы обслуживания с групповым поступлением заявок могут использоваться для моделирования процессов в сетях с пакетной коммутацией двумя способами: моделируя поступление групп пакетов одинаковых размеров или моделируя поступление отдельных пакетов разных размеров.

В этой главе рассматриваем многоканальную систему обслуживания с групповым поступлением заявок (пакетов) без ограничений на длину очереди с общей функцией отбрасывания пакетов. Изучаемая n -канальная система может служить моделью узла сети, в который поступают группы пакетов одинаковых размеров (одна заявка – один пакет). В то же время соответствующая одноканальная система (частный случай многоканальной системы при $n = 1$) может рассматриваться как модель узла с поступлением отдельных пакетов разных размеров [31].

Вместе с тем, отдельно изучаем одноканальную систему с ограниченным буфером, которая является моделью узла сети с поступлением отдельных пакетов разных размеров.

В этой главе ограничиваемся рассмотрением систем обслуживания с показательными распределениями промежутков времени между моментами поступления групп заявок. Это позволяет получить простой алгоритм определения стационарных характеристик не только для одноканальной, но и для многоканальной системы, в которой реализован механизм функции отбрасывания пакетов [32].

3.1 Одноканальная система с ограниченным буфером как модель поступления пакетов разных размеров

Изучаемая в этой главе система обслуживания может служить моделью узла сети в случае поступления отдельных пакетов разных размеров. Предполагается, что вероятность отбрасывания зависит от длины очереди в момент поступления пакета [33].

Рассмотрим одноканальную систему обслуживания $M^x/M/1/m$ с ограниченным буфером объема m . Промежутки времени между моментами поступления групп заявок входящего потока – независимые случайные величины, распределенные по показательному закону с параметром λ . Время обслуживания каждой заявки распределено по показательному закону с

параметром μ . С вероятностью α_k число заявок в группе входящего потока равно k ($k \geq 1$), причем

$$\sum_{k=1}^{\infty} a_k = 1, \quad \alpha_{(1)} = \sum_{k=1}^{\infty} k a_k < \infty. \quad (3.1)$$

Переходя к интерпретации этой системы обслуживания как модели узла сети для передачи пакетов данных, можно, например, предположить, что объем буфера m выразится в гигабайтах. Тогда с вероятностью a_k размер поступающего пакета равен k гигабайт.

Однако такая интерпретация имеет некоторые важные последствия. Например, если в буфере осталось 100 GB свободного места, то на обслуживание не может быть принят пакет размером 101 GB и больше. Поскольку пакеты нельзя делить, то нельзя принять в буфер фрагмент пакета.

Исходя из вышесказанного, будем использовать следующее правило принятия группы заявок в систему: прибывающая группа заявок получает отказ, если для всей группы не хватает места в буфере, даже если буфер еще полностью не заполнен.

В качестве алгоритма управления очередью применим общую функцию отбрасывания пакетов, определяемую набором вероятности β_i : Если в момент прибытия группы заявок размером k в системе находится i заявок и выполнено условие $i + k \leq m + 1$, то эта группа заявок принимается на обслуживание с вероятностью β_i ($0 \leq i \leq m$).

3.2 Стационарные характеристики одноканальной системы

Пусть $\rho_i(t)$ – вероятность того, что в момент времени t в системе находится i заявок ($i + k \leq m + 1$). Пределы $\rho_i = \lim_{t \rightarrow \infty} \rho_i(t)$ ($0 \leq i \leq m + 1$), очевидно, существуют. С целью упрощения системы уравнений для стационарных вероятностей ρ_i введем обозначения: $v_i = \mu/\beta_i$ ($1 \leq i \leq m$), $v_{m+1} = \mu$; $\alpha_i = \lambda/v_i$ ($1 \leq i \leq m + 1$); $\pi_i = \beta_i \rho_i / \rho_0$ ($1 \leq i \leq m$), $\pi_0 = 1$, $\pi_{m+1} = \rho_{m+1} / \rho_0$; $\alpha = \lambda/\mu$. Сделаем естественное предложение, что $\beta_0 = 1$.

Систему уравнений для стационарных вероятностей запишем как систему относительно π_i :

$$\begin{aligned} -\lambda \sum_{k=1}^{m+1} a_k + v_1 \pi_1 &= 0; \\ -\lambda \pi_i \sum_{k=1}^{m+1-i} a_k + \lambda \sum_{k=0}^{i-1} \pi_k \alpha_{i-k} - v_i \pi_i + v_{i+1} \pi_{i+1} &= 0, \quad (1 \leq i \leq m); \\ \rho_0 \left(1 + \sum_{k=1}^m \frac{\pi_k}{\beta_k} + \pi_{m+1} \right) &= 1. \end{aligned} \quad (3.2)$$

После решения системы уравнений (3.2) приходим к утверждению.

Теорема 1. Стационарные вероятности ρ_i определяется в виде

$$\rho_0 = \frac{1}{1 + \sum_{k=1}^m \frac{\pi_k}{\beta_k} + \pi_{m+1}}; \rho_i = \frac{\rho_0 \pi_i}{\beta_i}, \quad 1 \leq i \leq m; \rho_{m+1} = \rho_0 \pi_{m+1} \quad (3.3)$$

где π_i находим с помощью рекуррентных соотношений

$$\pi_0 = 1; \pi_i = a_i \sum_{j=0}^{i-1} \pi_j \sum_{k=1}^{m+1} \alpha_{k-j}, \quad 1 \leq i \leq m + 1. \quad (3.4)$$

Доказательство. Равенство (3.3) для ρ_0 следует из последнего уравнения (3.2) (условия нормировки). Рекуррентные соотношения (3.4) легко доказать методом математической индукции.

Приведем формулы для стационарных характеристик системы: средней длины очереди $E(Q)$, среднего числа заявок в системе $E(N)$, вероятности обслуживания пакета P_{sv} и среднего времени ожидания начала обслуживания $E(W)$:

$$\begin{aligned} E(Q) &= \sum_{k=1}^m k p_{k+1}; \quad E(N) = \sum_{k=1}^{m+1} k p_k = E(Q) + 1 - \rho_0; \\ P_{sv} &= \frac{1 - \rho_0}{\alpha \alpha_{(1)}}; \quad E(W) = \frac{E(Q)}{\lambda \alpha_{(1)} P_{sv}}. \end{aligned} \quad (3.5)$$

Отметим, что формула для P_{sv} получена как отношение стационарного среднего числа обслуженных пакетов за единицу времени $\mu(1 - \rho_0)/\alpha_{(1)}$ к стационарному среднему числу поступивших пакетов за единицу времени λ .

3.3 Одноканальная система с ограниченным буфером

Приведем результаты расчетов стационарных характеристик одноканальной системы, являющейся моделью узла сети в случае поступления отдельных пакетов разных размеров.

Предложим, что в одноканальный узел в среднем поступает 1,5 пакета информации за единицу времени (*е.в.*). В среднем за 1 *е.в.* непрерывно работающий узел обрабатывает 100 GB информации. Полная вместимость буфера – 400 GB (с учетом обрабатываемой в канале информации). Поступают пакеты трех видов в зависимости от их размеров. С вероятностью $a_1 = 0,2$ прибывший пакет содержит 100 GB, с вероятностью $a_2 = 0,3$ – 200 GB и с вероятностью $a_3 = 0,5$ – 300 GB.

Взяв за единицу измерения объем информации 100 GB и перейдя на язык теории массового обслуживания, получаем следующие значения параметров: $m = 3; \lambda = 1,5; \mu = 1; a_1 = 0,2; a_2 = 0,3; a_3 = 0,5; a_k = 0 (k \geq 4)$.

В качестве алгоритма управления очередью применим общую функцию отбрасывания пакетов, определяемую набором вероятностей $\beta_i (0 \leq i \leq 3)$. Рассмотрим 3 случая с разными набором вероятностей β_i :

$\beta_i = 1, (0 \leq i \leq 3)$ (случай, I, вероятностное отбрасывание пакетов не применяются);

$\beta_0 = \beta_1 = 1, \beta_2 = 0,8, \beta_3 = 0,5$ (случай II)

$\beta_0 = \beta_1 = 1, \beta_2 = \beta_3 = 0,1$ (случай III)

Стационарные вероятности $\rho_k (0 \leq k \leq 4)$ наличия в система k заявок стационарные характеристики $P_{sv}, E(Q), E(W)$ и $E(N)$, вычисленные для различных наборов β_i , приведены в таблице 3.1. Анализируя полученные результаты, видим, что применение вероятностного отбрасывания пакетов приводит к незначительному уменьшению вероятности обслуживания (на 3,6%, если сравнивать случаи III и I) и в то же время к более ощутимому улучшению характеристик очереди: $E(Q)$ уменьшается на 24,6 %, $E(W)$ на – 21,8 %, а $E(N)$ – на 17,2 %.

В таблице 3.2 приведено сравнение стационарных характеристик системы для случая II, вычисленных по формулам, предложенным в п. 3.2, с результатами имитационного моделирования, полученные с помощью системы GPSS World [34] (время моделирования $t = 10^6$) для различных распределений времени обслуживания: показательного распределения с параметром $\mu = 1$, равномерного распределения на отрезке $[0;2]$, равномерного распределения на отрезке $[0,5;1,5]$ и детерминированного значения, равного 1. Анализ данных таблиц показывает, что для показательного распределения времени обслуживания полученные аналитическим методом значения подтверждаются результатами имитационного моделирования. Уменьшение дисперсии распределения времени обслуживания приводит к незначительному улучшению показателей системы обслуживания: к увеличению вероятности обслуживания на 3,9 % и уменьшению $E(Q)$ и $E(W)$ соответственно на 7,7% и на 11% (если сравнивать детерминированное и показательное распределения времени обслуживания).

Таблица 3.1 - Стационарные характеристики системы для различных наборов β_i

№ набор а β_i	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	P_{sv}	$E(Q)$	$E(W)$	$E(N)$
I	0,065	0,098	0,225	0,335	0,275	0,270	1,723	1,844	2,658
	3	0	4	6	6	9	6	0	2
II	0,073	0,110	0,253	0,339	0,224	0,268	1,604	1,731	2,531
	4	1	1	0	5	6	5	6	2
III	0,098	0,148	0,340	0,277	0,134	0,261	1,300	1,442	2,201
	8	2	8	5	8	2	1	6	3

Таблица 3.2 - Сравнение стационарных характеристик системы для случая II с результатами имитационного моделирования для различных распределений времени обслуживания.

Вид распределения	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	P_{sv}	$E(Q)$	$E(W)$
Показательное (аналитич. метод)	0,073 4	0,110 1	0,253 1	0,339 0	0,224 5	0,268 6	1,604 5	1,731 6
Показательное (GPPS World)	0,072 9	0,110 4	0,252 9	0,338 5	0,225 4	0,269 0	1,606 0	1,733 0
Равномерное на [0;2] (GPPS World)	0,056 2	0,121 2	0,291 2	0,349 4	0,182 1	0,274 0	1,536 0	1,628 0
Равномерное на [0,5;1,5] (GPPS World)	0,043 0	0,132 0	0,314 3	0,352 0	0,158 7	0,278 0	1,495 0	1,562 0
Детерминированно $e=1$ (GPSS World)	0,039 1	0,134 4	0,321 8	0,353 4	0,151 4	0,279 0	1,481 0	1,541 0

3.4 Многоканальная система без ограничений на длину очереди

Изучаемая в этой главе n -канальная система обслуживания $M^x/M/n$ может служить моделью узла сети в случае поступления групп пакетов одинаковых размеров. Превышение объема буфера над поступающей нагрузкой считаем достаточным для того, чтобы не возникало необходимости отбрасывания пакетов по причине недостатка для них места в буфере.

Для рассмотренной выше одноканальной системы предполагалось, что «одна прибывающая группа заявок = одному поступающему пакету информации». Для данной системы считаем, что «одна прибывающая группа заявок = одной поступающей группе пакетов информации то есть «одна заявка = одному пакету».

Предположения п.3.1 относительно распределений для входящего потока и времени обслуживания заявки оставим без изменений, сделав дополнительное предположение, что

$$\alpha_{(1)} = \sum_{k=1}^{\infty} k^2 a_k < \infty. \quad (3.5)$$

В качестве алгоритма управления очередью, как и выше, применяем общую функцию отбрасывания пакетов, определяемую набором вероятностей β_i . Зафиксировав натуральное число h ($h > n$), определим набор вероятностей β_i с помощью равенств

$$\beta_i = \begin{cases} \beta_i & 0 \leq i \leq h; \\ \beta, & i \geq h + 1. \end{cases} \quad (3.6)$$

Если в момент прибытия группы заявок в системе находится i заявок, то эта группа заявок принимается на обслуживание с вероятностью β_i ($i \geq 0$). Предположим, что $\beta \leq \beta_i$, ($0 \leq i \leq h$).

Уточним введенные в п. 3.1 обозначения:

$$\begin{aligned} v_i &= \frac{\mu}{\beta_i}, \quad \alpha_i = \frac{\lambda}{v_i}, \quad i \geq 1; \\ \pi_i &= \frac{\beta_i \rho_i}{\rho_0}, \quad i \geq 0; \quad \alpha = \frac{\lambda}{\mu}, \quad \rho = \alpha \beta \alpha_{(1)}. \end{aligned} \quad (3.7)$$

Введем нумерацию состояний системы: s_0 – в системе нет заявок; s_i ($i \geq 1$) – в системе находится i заявок. Пусть $\rho_i(t)$ – вероятность того, что в момент времени t система находится в состоянии s_i . Предполагая, что существуют пределы $\rho_i = \lim_{t \rightarrow \infty} \rho_i(t)$ ($i \geq 0$) (условия их существования мы получим ниже), запишем систему уравнений для определения π_i

$$\begin{aligned} -\lambda \pi_0 + v_1 \pi_1 &= 0; \\ -(\lambda + i v_i) \pi_i + \lambda \sum_{k=0}^{i-1} \pi_k \alpha_{i-k} + (i+1) v_{i+1} \pi_{i+1} &= 0, \quad 1 \leq i \leq n-1; \\ -(\lambda + n v_i) \pi_i + \lambda \sum_{k=0}^{i-1} \pi_k \alpha_{i-k} + n v_{i+1} \pi_{i+1} &= 0, \quad i \leq n; \\ \rho_0 \sum_{k=0}^{\infty} \frac{\pi_k}{\beta_k} &= 1. \end{aligned} \quad (3.8)$$

Введем обозначения:

$$A_i = 1 - \sum_{k=1}^i \alpha_k, \quad i \geq 0, \quad A_0 = 1.$$

Теорема 2. Если

$$\alpha_{(1)} < \infty; \quad \rho < n; \quad \beta \leq \beta_i \quad (0 \leq i \leq h), \quad (3.9)$$

то стационарные вероятности ρ_i ($i \geq 0$) существуют и определяются в виде

$$\rho_0 = \frac{n-p}{n + \sum_{k=1}^{n-1} \frac{(n-k)\pi_k}{\beta_k} + \alpha \alpha_{(1)} (\beta_0 - \beta + \sum_{k=1}^h (1 - \frac{\beta}{\beta_k}) \pi_k)}; \quad \rho_i = \frac{\rho_0 \pi_i}{\beta_i}, \quad i \geq 1, \quad (3.10)$$

где π_i находим с помощью рекуррентных соотношений

$$\pi_0 \beta_0; \quad \pi_{i+1} = \frac{\alpha_{i+1}}{i+1} \sum_{k=0}^i A_{i-k} \pi_k, \quad 0 \leq i \leq n-1 \quad (3.11)$$

$$\pi_{i+1} = \frac{\alpha_i + 1}{n} \sum_{k=0}^i A_{i-k} \pi_k, \quad i \geq n$$

Доказательство. Пользуясь уравнениями (3.8), рекуррентные соотношения (3.11) легко доказать методом математической индукции. Для доказательства первого равенства (3.10) с помощью условия нормировки (3.8) находим

$$\begin{aligned} \sum_{i=0}^{\infty} \sum_{k=0}^i A_{i-k} \pi_k &= \sum_{i=0}^{\infty} \pi_i \sum_{k=0}^{\infty} A_k = \frac{\alpha_{(1)}}{\rho_0} \sum_{i=0}^{\infty} \beta_i \rho_i = \frac{\alpha_{(1)}}{\rho_0} \left(\beta + \sum_{i=0}^h (\beta_i - \beta) \rho_i \right); \\ \sum_{i=1}^n \frac{i \pi_i}{\alpha_i} + n \sum_{i=n+1}^{\infty} \frac{\pi_i}{\alpha_i} &= \frac{1}{\alpha} \left(\sum_{i=1}^n \frac{i \pi_i}{\beta_i} + n \left(\frac{1}{\rho_0} - \sum_{i=0}^n \frac{\pi_i}{\beta_i} \right) \right). \end{aligned} \quad (3.12)$$

Учитывая соотношения (3.11), получаем равенство

$$\alpha \alpha_{(1)} \left(\frac{\beta}{\rho_0} + \beta_0 - \beta + \sum_{i=0}^h (\beta_i - \beta) \frac{\pi_i}{\beta_i} \right) = \sum_{i=1}^n \frac{i \pi_i}{\beta_i} + n \left(\frac{1}{\rho_0} - 1 - \sum_{i=0}^n \frac{\pi_i}{\beta_i} \right) \quad (3.13)$$

приводящее к выражению (3.10) для ρ_0 .

Итак, при выполнении условий (3.9) система уравнений равновесия (3.8) имеет нетривиальное решение. Пусть $\xi(t)$ – число заявок в системе в момент времени t . Тогда $\{\xi(t), |t \geq 0\}$ – случайный процесс с дискретными состояниями $s_i (i \geq 0)$. Поскольку все состояния этого неприводимого марковского процесса сообщаются, то из критерия регулярности [35] следует его регулярность, а из эргодической теоремы Фостера вытекает, что условия теоремы 2 является достаточными для существования стационарных вероятностей $\rho_i (i \geq 0)$. Теорема доказана.

3.5 Стационарные характеристики многоканальной системы

Обозначим через $N(F)(z)$ и $D(F)(z)$ соответственно числитель и знаменатель в выражении для некоторой функции $(F)(z)$. Рассмотрим производящие функции

$$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^i; \quad A(z) = \sum_{i=0}^{\infty} \alpha_i z^i. \quad (3.14)$$

Теорема 3. Производящая функция $\pi(z)$ определяется в виде

$$\begin{aligned} \pi(z) &= \frac{N(\pi)(z)}{D(\pi)(z)}; \quad D(\pi)(z) = n\mu(1-z) - \lambda\beta_z(1-A(z)); \\ N(\pi)(z) &= \mu(1-z) \left(n(\beta_0 + \sum_{i=1}^h \pi_i z^i + \beta \sum_{i=n}^h \frac{\pi_i z^i}{\beta_i}) - \beta \sum_{i=1}^{n-1} \frac{i \pi_i z^i}{\beta_i} \right). \end{aligned} \quad (3.15)$$

Доказательство. Умножая i -е уравнение системы (3.8) на z^i ($i \geq 0$) и суммируя, с учетом равенства

$$\sum_{i=n}^{\infty} v_i \pi_i z^i = \mu \left(\sum_{i=n}^h \frac{\pi_i z^i}{\beta_i} + \frac{1}{\beta} (\pi(z) - \sum_{i=0}^h \pi_i z^i) \right) \quad (3.16)$$

получаем

$$\begin{aligned} & -\lambda \left(1 - A(z) \right) \pi(z) - \frac{n\mu}{\beta} \left(1 - \frac{1}{z} \right) \pi(z) \\ & - n\mu \left(1 - \frac{1}{z} \right) \left(\sum_{i=n}^h \frac{\pi_i z^i}{\beta_i} - \frac{1}{\beta} \sum_{i=0}^h \pi_i z^i \right) - \left(1 - \frac{1}{z} \right) \sum_{i=1}^{n-1} i v_i \pi_i z^i = 0. \end{aligned} \quad (3.17)$$

Отсюда находим $\pi(z)$ в виде (3.15). Теорема доказана.

Лемма 1. Производная функции $\pi(z)$ при $z = 1$ вычисляется по формуле:

$$\begin{aligned} \pi'(1) = \frac{1}{2(n-p)^2} & \left(2(n-p) \left(n \sum_{i=1}^h i \pi_i - n\beta \sum_{i=n}^h \frac{i \pi_i}{\beta_i} - \beta \sum_{i=1}^{n-1} \frac{i^2 \pi_i}{\beta_i} \right) + \right. \\ & \left. \alpha\beta (\alpha_{(1)} + \alpha_{(2)}) \left(n \sum_{i=0}^h \pi_i - n\beta \sum_{i=n}^h \frac{\pi_i}{\beta_i} - \beta \sum_{i=1}^{n-1} \frac{i \pi_i}{\beta_i} \right) \right) \end{aligned} \quad (3.18)$$

Доказательство. При вычислении производной $\pi'(1)$ будем учитывать равенства

$$A'(1) = \sum_{k=1}^{\infty} k \alpha_k = a_{(1)}; \quad A''(1) = \sum_{k=2}^{\infty} k(k-1) \alpha_k = a_{(2)} - a_{(1)} \quad (3.19)$$

Поскольку $A(1) = \sum_{k=1}^{\infty} \alpha_k = 1$, то для функции $\pi(z)$, определяемой в виде (3.15), выполняется соотношения

$$\begin{aligned} N(\pi)(1) = D(\pi)(1) = N(\pi')(1) = D(\pi')(1) = N'(\pi')(1) = D'(\pi')(1) = 0; \\ D''(\pi')(1) \neq 0. \end{aligned} \quad (3.20)$$

Из (3.20) следует равенство

$$\pi'(1) = \frac{N''(\pi')(1)}{D''(\pi')(1)}. \quad (3.21)$$

Поскольку

$$\begin{aligned} N(\pi')(z) &= N'(\pi)(z) \cdot D(\pi)(z) - N(\pi)(z) \cdot D'(\pi)(z); \\ N'(\pi')(z) &= N''(\pi)(z) \cdot D(\pi)(z) - N(\pi)(z) \cdot D''(\pi)(z); \end{aligned}$$

$$N''(\pi')(z) = N'''(\pi)(z) \cdot D(\pi)(z) - N''(\pi)(z) \cdot D'(\pi)(z); \quad (3.22)$$

$$-N'(\pi)(z) \cdot D''(\pi)(z) - N(\pi)(z) \cdot D'''(\pi)(z),$$

то с учетом (3.20) получим формулу

$$N''(\pi')(1) = N''(\pi)(1) \cdot D'(\pi)(1) - N'(\pi)(1) \cdot D''(\pi)(1), \quad (3.23)$$

пользуясь которой с помощью (3.21) приходим к равенству (3.18). Лемма доказана.

Используя стационарное распределение числа заявок и производную $\pi'(1)$, можно получить простые формулы для основных стационарных характеристик системы.

Учитывая, что

$$\pi'(1) = \sum_{k=1}^{\infty} k\pi_k = \sum_{k=1}^h k\pi_k + \frac{\beta}{\rho_0} \sum_{k=h+1}^{\infty} k\rho_k,$$

$$\sum_{k=h+1}^{\infty} k\rho_k = \frac{\rho_0}{\beta} (\pi'(1) - \sum_{k=1}^h k\pi_k), \quad (3.24)$$

формулы для стационарной средней длины очереди в системе и стационарного среднего числа заявок (пакетов) в системе

$$E(Q) = \sum_{k=n+1}^{\infty} (k-n)\rho_k, \quad E(N) = \sum_{k=1}^{\infty} k\rho_k, \quad (3.25)$$

приводим к виду

$$E(Q) = \sum_{k=n+1}^h k\rho_k - n(1 - \sum_{k=0}^n \rho_k) + \frac{\rho_0}{\beta} (\pi'(1) - \sum_{k=1}^h k\pi_k), \quad (3.26)$$

$$E(N) = \sum_{k=1}^h k\rho_k + \frac{\rho_0}{\beta} (\pi'(1) - \sum_{k=1}^h k\pi_k). \quad (3.27)$$

Формула для стационарного среднего числа обслуженных заявок (пакетов) за единицу времени после использования условия нормировки принимает вид

$$\overline{N_{SV}} = \mu(\sum_{k=1}^{n-1} k\rho_k + n \sum_{k=n}^{\infty} \rho_k) = \mu(n(1 - \rho_0) - \sum_{k=1}^{n-1} (n-k)\rho_k). \quad (3.28)$$

Стационарная вероятность обслуживания заявки (пакета) P_{sv} и стационарное среднее время ожидания начала обслуживания $E(W)$ определяется в виде

$$P_{sv} = \frac{\overline{N_{SV}}}{\lambda\alpha_{(1)}} = \frac{n(1-\rho_0) - \sum_{k=1}^{n-1} (n-k)\rho_k}{\alpha\alpha_{(1)}}, \quad E(W) = \frac{E(Q)}{N_{SV}}. \quad (3.29)$$

3.6 Случай ординарного входящего потока

Поскольку для ординарного потока заявок выполняются равенства

$$\alpha_1 = 1; a_k = 0 (k \geq 2); A_k = 0 (k \geq 1), \quad (3.30)$$

то рекуррентные соотношения (3.11) упрощаются, что позволяет получить выражения для $\pi_k (k \geq 0)$ в явном виде

$$\begin{aligned} \pi_k &= \frac{\alpha^k}{k!} \sum_{i=0}^k \beta_i, \quad 0 \leq k \leq n; \quad \pi_k = \frac{\alpha^k}{n! n^{k-n}} \prod_{i=0}^k \beta_i, \quad n+1 \leq k \leq h; \\ \pi_k &= \frac{\alpha^k \beta^{k-h}}{n! n^{k-n}} \prod_{i=0}^h \beta_i, \quad k \geq h+1. \end{aligned} \quad (3.31)$$

Стационарное распределение числа заявок в системе существует при выполнении условия $\alpha\beta < n$ и определяется равенствами

$$\begin{aligned} \rho_0 &= \left(1 + \sum_{k=1}^n \frac{\alpha^k}{k!} \prod_{i=0}^{k-1} \beta_i + \frac{1}{n!} \sum_{k=n+1}^h \frac{\alpha^k}{n^{k-n}} \prod_{i=0}^{k-1} \beta_i + \frac{\alpha^{h+1}}{n! n^{h-n} (n - \alpha\beta)} \prod_{i=0}^h \beta_i \right)^{-1}; \\ \rho_k &= \frac{\rho_0 \pi_k}{\beta_k}, \quad k \geq 1. \end{aligned} \quad (3.32)$$

Для стационарных характеристик системы получаем формулы

$$\begin{aligned} E(Q) &= \sum_{k=n+1}^h (k-n) \rho_k + \frac{\rho_0 \alpha^{h+1} (n + (h-n)(n - \alpha\beta))}{n! n^{h-n} (n - \alpha\beta)^2} \prod_{i=0}^h \beta_i; \quad E(W) = \frac{E(Q)}{\lambda P_{sv}}; \\ P_{sv} &= \frac{1}{\alpha} \left(\sum_{k=1}^{n-1} k \rho_k + n \sum_{k=n}^h \rho_k + \frac{\rho_0 \alpha^{h+1}}{n! n^{h-n} (n - \alpha\beta)} \prod_{i=0}^h \beta_i \right). \end{aligned} \quad (3.33)$$

3.7 Многоканальная система без ограничений на длину очереди

Приведем результаты расчетов стационарных характеристик многоканальной системы, выполненных с помощью аналитических соотношений, полученных в п.3.5.

Вычисления проводились для трехканальной системы обслуживания ($n = 3$), в которую заявки (пакеты информации одинаковых размеров) прибывают группами численностью от одной до трех с распределением числа заявок в группе, задаваемым вероятностями: $a_1 = 0,2$; $a_2 = 0,3$; $a_3 = 0,5$; $a_k = 0 (k \geq 4)$. В качестве алгоритма управления очередью применим общую функцию отбрасывания групп заявок с пороговым значением $h = 5$. Рассмотрим 4 случая с разными наборами вероятности β_i :

$\beta_i = 1 (0 \leq i \leq 3); \beta_4 = 0,8; \beta_5 = 0,4; \beta_i = \beta = 0,1 (i \geq 6)$ (случай I);
 $\beta_i = 1 (0 \leq i \leq 3); \beta_4 = 0,8; \beta_5 = 0,4; \beta_i = \beta = 0,2 (i \geq 6)$ (случай II);
 $\beta_i = 1 (0 \leq i \leq 3); \beta_4 = 0,8; \beta_5 = 0,4; \beta_i = \beta = 0,3 (i \geq 6)$ (случай III);
 $\beta_i = 1 (0 \leq i \leq 3); \beta_4 = 0,8; \beta_5 = 0,4; \beta_i = \beta = 0,4 (i \geq 6)$ (случай IV).

Стационарные вероятности $\rho (0 \leq k \leq 8)$ наличия в системе k заявок и стационарные характеристики $P_{sv}, E(Q)$ и $E(W)$, вычисленные для различных наборов β_i , приведены в таблице 3.3. Анализируя полученные результаты, видим, что увеличение вероятности β в 4 раза (если сравнивать случаи IV и I) приводит к незначительному увеличению вероятности обслуживания (на 3,9%) и в тоже время к более ощутимому ухудшению характеристик очереди: $E(Q)$ увеличивается в 8,2 раза, а $E(W)$ – 7,9 раза.

В таблице 3.4 приведено сравнение стационарных характеристик системы для случая II, вычисленных по формулам, предложенным в п. 3.5, с результатами имитационного моделирования, полученными с помощью системы GPSS World (время моделирования $t = 10^6$) для различных распределений времени обслуживания: показательного распределения с параметром $\mu = 1$, равномерного распределения на отрезке $[0;2]$, равномерного распределения на отрезке $[0,5;1,5]$ и детерминированного значения, равного 1. Анализ данных таблицы показывает, что для показательного распределения времени обслуживания полученных аналитическим методом значения подтверждаются результатами имитационного моделирования. Уменьшение дисперсии распределения времени обслуживания приводит к незначительному улучшению показателей системы обслуживания: к увеличению вероятности обслуживания на 3,8% и уменьшению $E(Q)$ и $E(W)$ соответственно на 9,9% и на 13,2% (если сравнивать детерминированное и показательное распределение времени обслуживания). Программа для GPSS World приведена в приложении.

Таблица 3.3 - Стационарные характеристики трехканальной системы для различных наборов β_i

№ набор ов β_i	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	P_{sv}	$E(Q)$	$E(W)$
I	0,00 95	0,02 85	0,05 41	0,08 16	0,13 90	0,20 35	0,21 12	0,14 19	0,07 18	0,41 46	2,51 79	0,88 02
II	0,07 9	0,02 360	0,04 48	0,06 76	0,11 53	0,16 87	0,17 51	0,13 51	0,08 88	0,41 80	3,34 81	1,16 08
III	0,00 55	0,01 66	0,03 15	0,04 75	0,08 08	0,11 85	0,12 29	0,10 71	0,08 53	0,42 30	5,57 20	1,90 90
IV	0,00 18	0,00 55	0,01 04	0,01 58	0,02 69	0,03 93	0,04 08	0,03 97	0,03 68	0,43 09	20,7 619	6,98 33

Таблица 3.4 - Сравнение стационарных характеристик трехканальной системы для случая II с результатами имитационного моделирования для различных распределений времени обслуживания

Вид распределения	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	P_{sv}	$E(Q)$	$E(W)$
Показатель (аналит.метод)	0,00 79	0,02 36	0,04 48	0,06 76	0,11 53	0,16 87	0,41 80	3,34 81	1,16 08
Показатель (GPSS World)	0,00 80	0,02 34	0,04 47	0,06 79	0,11 54	0,16 88	0,41 80	3,34 70	1,16 10
Равномерн. на [0; 2] (GPSS World)	0,00 04	0,00 41	0,01 83	0,05 38	0,12 82	0,21 05	0,43 10	3,15 30	1,06 00
Равномерн. на [0,5; 1,5] (GPSS World)	0,00 01	0,00 12	0,01 14	0,05 19	0,13 65	0,22 59	0,43 30	3,06 30	1,02 50
Детерминированное=1 (GPSS World)	0,00 02	0,00 63	0,04 75	0,14 48	0,23 69	0,23 23	0,43 40	3,01 60	1,00 80

3.8 Разработка модели алгоритма управления очередями с учетом повторных поступлением отброшенных пакетов

В рассмотренных нами алгоритмах управления очередями не учитывается повторное поступление в узел отброшенных пакетов. Известно, что протоколы, работающие с установлением соединения, в случае если пакет отброшен и не доставлен получателю организуют повторную передачу пакета. Отправитель нумерует пакеты и ожидает от приемника служебное сообщение о том, что пакет получен. Время ожидания служебного извещения ограничено, если оно не получено за определенный промежуток времени, то пакет считается утерянным, и его отправляют повторно. Такой процесс называется квитированием, он является одним из методов обеспечения надежной связи [38].

Основной задачей управления очередями является минимизация средней длины очереди и справедливое распределение объема буфера между различными потоками данных. Алгоритмы управления очередями различаются в основном правилом по которому отбрасываются пакеты. В самом простом случае пакеты отбрасываются при достижении некоторого порога.

Основа RED алгоритма это – вычисление среднего размера очереди и вычисления вероятности отбрасывания пакетов. При вычислении вероятности отбрасывания пакетов алгоритм RED вычисляет не текущий, а экспоненциально взвешенный средний размер очереди. Текущий средний размер очереди определяется на основании текущего действительного и предыдущего среднего размера. Использование механизмом RED среднего размера очереди обусловлено стремлением реагировать только на продолжительную перегрузку сети и «не замечать» моментальных всплесков трафика. Средний размер очереди avg

вычисляется по формуле:

$$avg = (1 - W) * avg_old + W * q, \quad (3.34)$$

где W – это весовой коэффициент очереди, $W = 2^{-\frac{1}{n}}$; n – экспоненциальный весовой коэффициент; q – текущий размер очереди; avg_old – значение средней длины очереди на предыдущем шаге измерений.

Вероятность отбрасывания пакетов p_a представляет собой функцию, линейно зависящую от среднего размера очереди (рисунок 3.1). Помимо этого, данная функция зависит также от минимального порогового значения min_th , максимального порогового значения max_th и знаменателя граничной вероятности max_p (mark probability denominator), определяющего часть отбрасываемых пакетов при достижении средним размером очереди максимального порогового значения. Вероятность отбрасывания пакетов рассчитывается по формуле:

$$P_a = \begin{cases} 0, & \text{если } avg < min_th \\ max_p * \frac{avg - min_th}{max_th - min_th}, & \text{если } min_th \leq avg < max_th \\ 1, & \text{если } max_th \leq avg \end{cases} \quad (3.35)$$

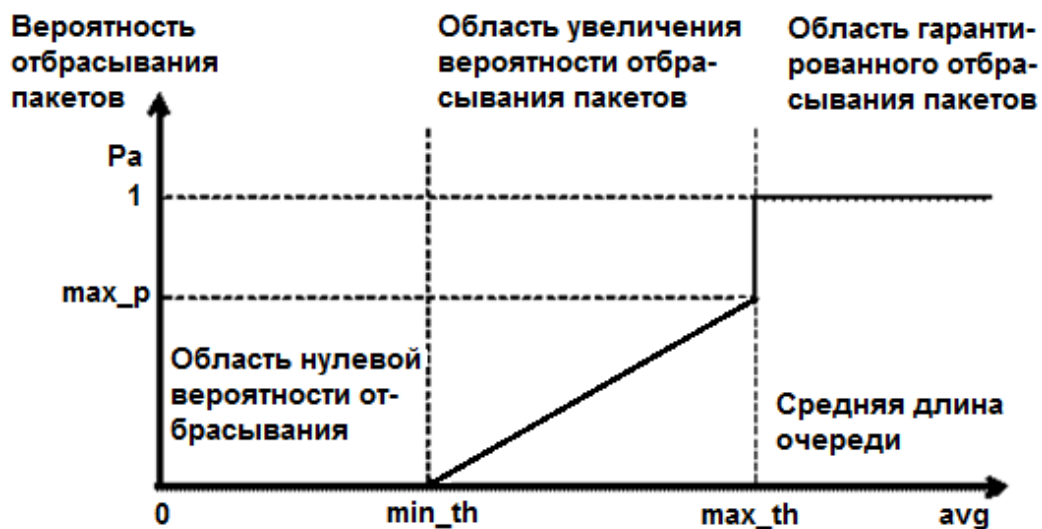


Рисунок 3.1 – Зависимость вероятности отбрасывания пакетов механизмом RED от средней длины очереди

Когда средний размер очереди превышает минимальное пороговое значение, механизм RED начинает отбрасывать пакеты. Интенсивность отбрасывания пакетов возрастает прямо пропорционально возрастанию среднего размера очереди до тех пор, пока он не достигнет максимального порогового значения. Когда средний размер очереди превышает

максимальное пороговое значение, механизм RED отбрасывает все пакеты, предназначенные для постановки в очередь.

На рисунке 3.2 показан алгоритм управления очередями в узле сети с коммутацией пакетов, а на рисунке 3.3 алгоритм с повторным поступлением в узел отброшенных пакетов.

Алгоритмы реализованы в виде программ, разработанных в среде имитационного моделирования GPSS World [34].

Узел коммутации в программах рассматривался как одноканальная СМО с ожиданием. Роль транзакта в модели играет пакет. Статистические данные о задержке в имитационной модели накапливались в виде значения числового атрибута M1, который характеризует время пребывания транзакта в модели, включающего в себя время ожидания в очереди (буфере) и времени обслуживания.

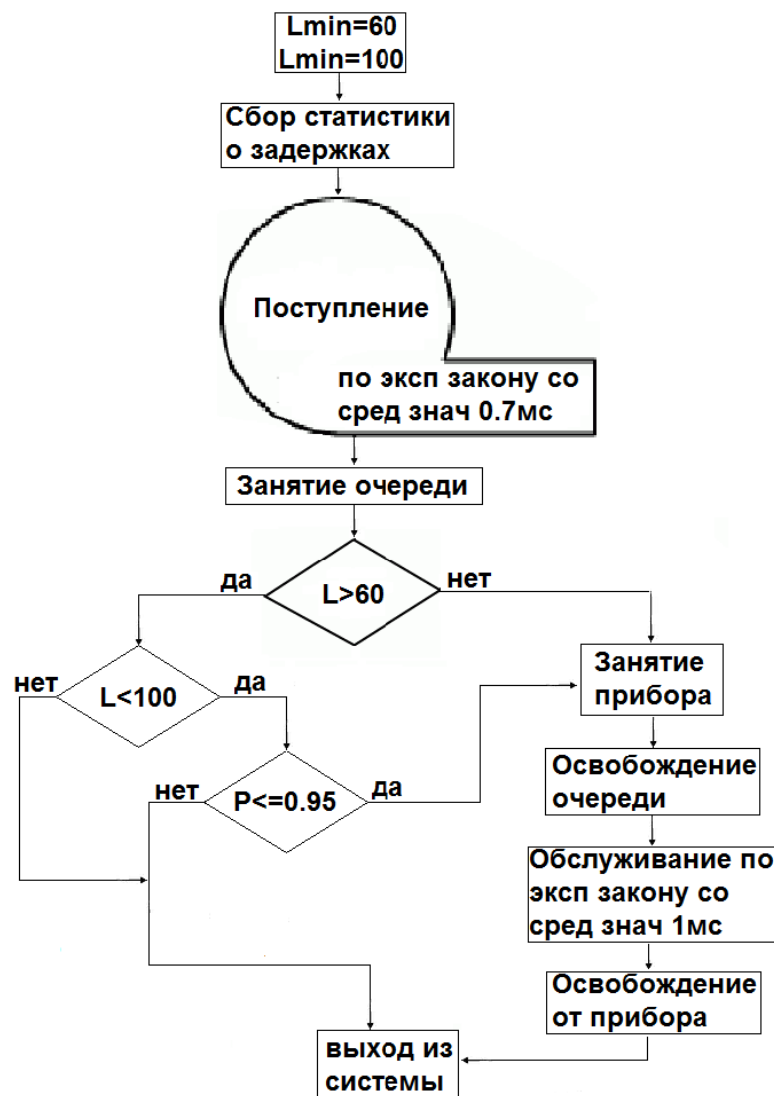


Рисунок 3.2 – Алгоритм управления очередями в узле сети с коммутацией пакетов без повторного поступления в узел отброшенных пакетов

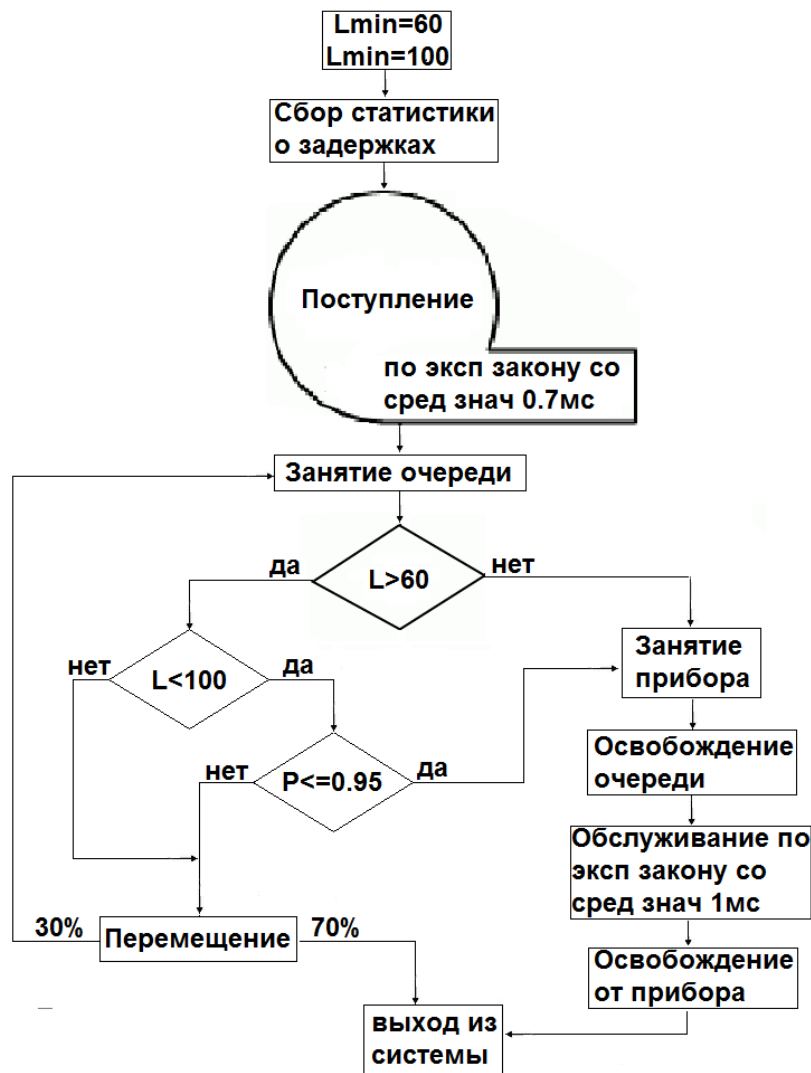


Рисунок 3.3 – Алгоритм управления очередями в узле сети с коммутацией пакетов с повторным поступлением в узел отброшенных пакетов.

На рисунке 3.4 представлена гистограмма, полученная в результате работы программы, реализующей первый алгоритм, показывающая распределение времени пребывания пакета в узле коммутации, математическое ожидание составляет $M=103$ мс, среднеквадратическое отклонение $\sigma=50,6$ мс.

На рисунке 3.5 представлена гистограмма, показывающая распределение времени пребывания пакета в узле коммутации, при реализации второго алгоритма, математическое ожидание составляет $M=106$ мс, среднеквадратическое отклонение $\sigma=64,3$ мс. Как видно из этих гисторграмм время пребывания пакетов в узле коммутации увеличивается.

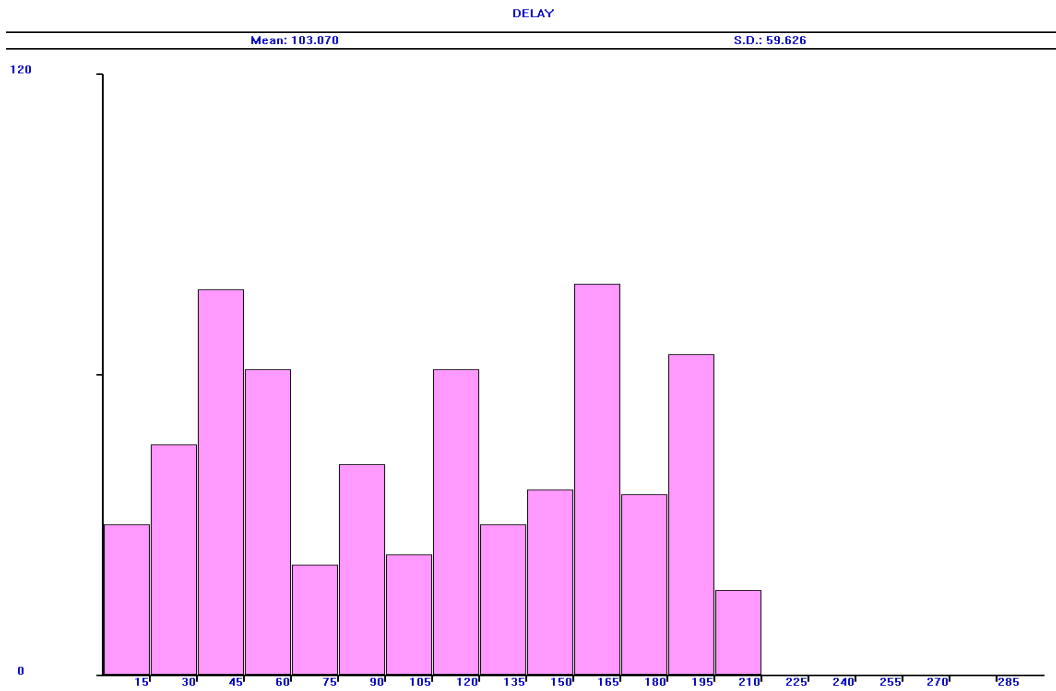


Рисунок 3.4 - Распределение времени пребывания пакета в узле коммутации в первом алгоритме

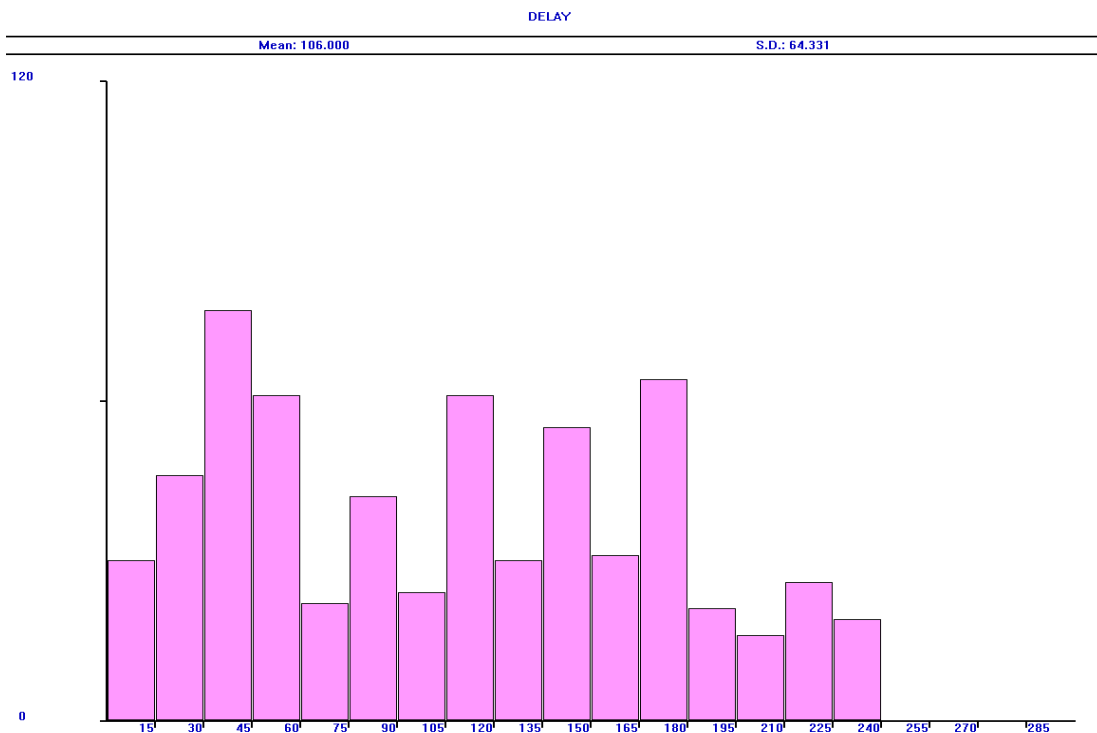


Рисунок 3.5 - Распределение времени пребывания пакета в узле коммутации во втором алгоритме

На рисунке 3.6 представлена зависимость между значением нагрузки системы ρ и значением $M1$. Значение нагрузки есть отношение:

$$\rho = \frac{\lambda}{\mu}, \quad (3.36)$$

где λ - интенсивность поступающего потока пакетов, μ - интенсивность обслуживания в узле коммутации.

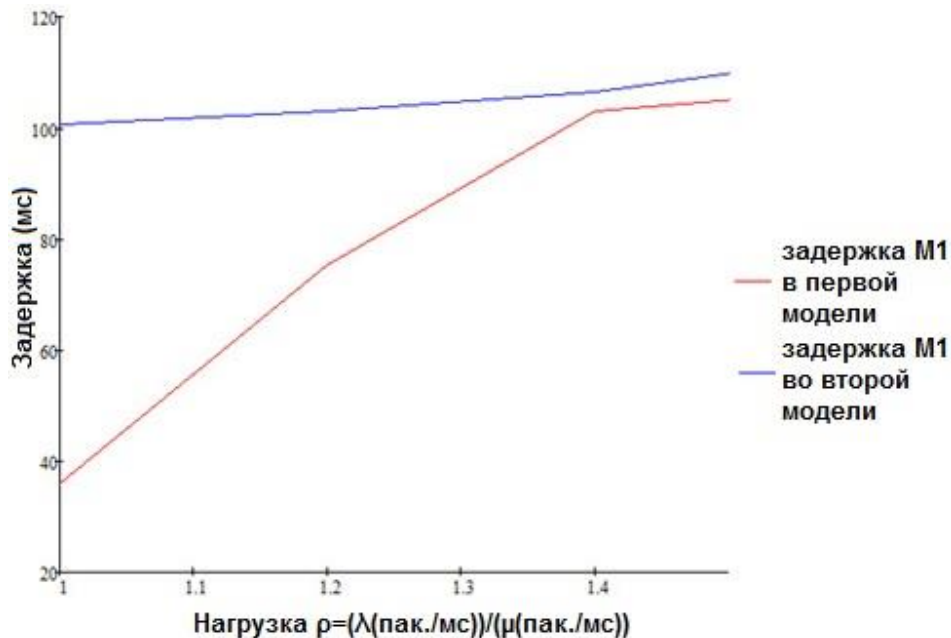


Рисунок 3.6 – График зависимости задержки M1 от нагрузки

На рисунке 3.6 представлен результат моделирования и видно, что в алгоритме с повторным поступлением в узел отброшенных пакетов задержка пакета значительно больше.

Сравнительный анализ алгоритма управления очередями в узле сети с коммутацией пакетов RED без учета отброшенных пакетов и алгоритма с учетом повторного поступления в очередь отброшенных пакетов показывает, что в алгоритме с повторным поступлением в узел отброшенных пакетов время пребывания пакета в узле значительно повышается. Полученный результат показывает, что при проектировании алгоритма необходимо учитывать поступление отброшенных пакетов в очередь, поскольку данный фактор влияет на важный показатель качества обслуживания – задержку пакета при передаче.

Заключение

В магистерской диссертации проведено исследование алгоритмов управления очередями в узле коммутации сети следующего поколения. Рассмотрены системы обслуживания, моделирующие работу узлов сетей передачи пакетов данных с использованием алгоритмов вероятностного отбрасывания пакетов. В системе GPSS World разработаны имитационные модели узла сети передачи пакетов данных с использованием алгоритма без повторного поступления в узел отброшенных пакетов и с их повторным поступлением.

В первой главе дается обзор методов поддержки качества обслуживания в сетях, построенных на базе IP-ориентированных протоколов. Представлены основные подходы к решению задачи обеспечения гарантированного качества обслуживания в IP-сетях, базирующиеся на моделях интегрированных и дифференцированных услуг. В заключение можно отметить следующее: решение задачи обеспечения требуемого качества обслуживания в сетях IP, безусловно, может быть достигнуто прямым путем – на основе предоставления гарантированной полосы пропускания, повышения производительности сетевых устройств – маршрутизаторов и шлюзов, использовании магистралей с высокими пропускными способностями. Однако, наиболее целесообразным представляется применение гибких методов, которые обеспечивают требуемые показатели качества обслуживания при эффективном использовании ресурсов сети для большого набора различных приложений, включая и наиболее критичные аудио- и видеоприложения реального времени.

Во второй главе исследование и определены особенности алгоритма RED и его модификаций, выявлены общие признаки модификаций RED и построена их классификация, которая априори является частичной, поскольку применённый фасетный метод не задаёт полной таксономии предметной области. Для иллюстрации построенной классификации приведено описание наиболее характерных для каждого класса модифицированных алгоритмов RED и составлена сводная таблица.

В третьей главе изучены системы обслуживания, моделирующие работу узлов сетей передачи пакетов данных с использованием алгоритмов вероятностного отбрасывания пакетов. Приведено сравнение результатов тестовых расчетов, выполненных с помощью полученных аналитических соотношений для случая показательного распределения времени обслуживания, с результатами имитационного моделирования для различных распределений времени обслуживания. Незначительное расхождение в значениях стационарных характеристик для различных распределений времени обслуживания свидетельствует о возможности использования полученных аналитических соотношений для проведения оценочных расчетов при разработке оптимальных алгоритмов активного управления очередью. Входящий поток моделируется с помощью обобщённого пуассоновского

процесса, а время обслуживания показательно распределено. Предложен алгоритм определения стационарного распределения числа заявок и стационарных характеристик системы (средней длины очереди, среднего времени ожидания в очереди, вероятности обслуживания заявок).

Сравнительный анализ алгоритма управления очередями в узле сети с коммутацией пакетов RED без учета отброшенных пакетов и алгоритма с учетом повторного поступления в очередь отброшенных пакетов показывает, что в алгоритме с повторным поступлением в узел отброшенных пакетов время пребывания пакета в узле значительно повышается. Полученный результат показывает, что при проектировании алгоритма необходимо учитывать поступление отброшенных пакетов в очередь, поскольку данный фактор влияет на важный показатель качества обслуживания – задержку пакета при передаче.

Список литературы

1. McDysan. QoS and Traffic Management in IP and ATM Networks // McGraw-Hill. 2000.
2. Е.А. Кучерявый. Управление трафиком и качество обслуживания в сети Интернет//СПб, Наука и Техника. 2004.
3. Р. Кох, ГГ. Яновский. Эволюция и конвергенция в электросвязи//М., Радио и связь. 2001.
4. МСЭ-Т Recommendation Y.1540. IP Packet Transfer and Availability Performance Parameters//December 2002.
5. МСЭ-Т Recommendation Y.1541. Network Performance Objectives for IP-Based Services//May 2002.
6. МСЭ-Т Recommendation Y.1291. An Architectural Framework for Support of Quality of Service in Packet Networks //May 2004.
7. L. Zhang, R. Braden. Resource reservation Protocol//RFC-2205, September 1997.
8. S. Blake et al. Architecture for Differentiated Services//RFC-2475, December 1998.
9. Floyd S., Jacobson V. Random Early Detection Gateways for Congestion Avoidance // IEEE/ACM Transactions on Networking, 1993, No 1(4). - P p . 397-413.
10. Floyd S., Gummadi R., Shenker S. Adaptive RED: An Algorithm for Increasing the Robustness of RED's Active Queue Management. - 2001.
11. Floyd S. Explicit Congestion Notification (ECN) Mechanism in the TCP/IP Protocol // ACM Computer Communications Review. - 1994. - Vol. 24.
12. Floyd S. RED: Discussions of Setting Parameters. - 1997.
13. Королькова А. В., Кулябов Д. С. Реализация модели схемы управления трафиком в IP-сетях с предоставлением дифференцированных услуг с использованием модуля DiffServ пакета NS-2 // Сборник докладов. IX-я научная конференция МГТУ “Станкин” и “Учебно-научного центра математического моделирования МГТУ “Станкин” - ИММ РАН” по математическому моделированию и информатике.— М.: “ЯНУС-К”, ИЦ ГОУ МГТУ “Станкин”, 2006. - С. 116-118.
14. Кулябов Д. С, Королькова А. В. Инструментальные средства моделирования DiffServ // Фундаментальные физико-математические проблемы и моделирование технико-технологических систем. Ежегодный сборник научных трудов / под ред. Л. А. Уваровой. - М.:Янус-К, 2008. - Вып. 11. - С. 161-167.
15. Que D., Chen Z., Chen B. An Improvement Algorithm Based on RED and Its Performance Analysis // IEEE. - 2008. - Pp. 2005-2008.
16. Techniques for Eliminating Packet Loss in Congested TCP/IP Network: Techrep CSE-TR-349-97/W.-C. Feng, D. Kandlur, D. Saha, K. Shin/U. Michigan. - 1997.

17. A Self-Configuring RED Gateway / W.-C. Feng, D. Kandlur, D. Saha, K. Shin // Infocom. - 1999.
18. Dynamics of Random Early Detection / D. Lin, D. Lin, R. Morris, R. Morris // Proceedings of ACM SIGCOMM. — Sophia Antipolis, France: 1997. - Pp. 127-137.
19. Kim T.-H., Lee K.-H. Refined Adaptive RED in TCP/IP Networks // SICE-ICASE International Joint Conference. — Bexco, Busan, Korea: 2006.
20. POWARED for Non-Linear Adaptive RED / B. K. Ng, M. S. Uddin, A. A. Y. M. Abusin, D. Chieng // Asia-Pacific Conference on Communications, Perth, Western Australia, 3 - 5 October 2005.
21. Zheng B., Atiquzzaman M. DSRED: A New Queue Management Scheme for the Next Generation Internet // IEICE TRANS. COMMUN.- 2006. - Vol. E89-B, No 3.
22. Lannaccone G., May M., Diot C. Aggregate Traffic Performance with Active Queue Management and Drop From Tail // SIGCOMM Comput. Commun. Rev. — 2001. — Vol. 31, No 3. — Pp. 4-13. — ISSN 0146-4833.
23. Ryoo I., Yang M. A State Dependent RED: An Enhanced Active Queue Management Scheme for Real-Time Internet Services // IEICE Trans. Commun. - 2006. - Vol. E89-B, No 2. - Pp. 614-617.
24. Ott T., Lakshman T., Wong L. SRED: Stabilized RED // Proceedings of INFOCOM'99. - 1999. - Pp. 1346-1355.
25. Clark D. D., Fang W. Explicit Allocation of Best-Effort Packet Delivery Service // IEEE/ACM Transactions on Networking. - 1998. - Vol. 6, No 4. - Pp. 362-373.
26. Wu H.-M., Wu C.-C, Lin W. SF-RED - a Novel Server-based AQM to Provide Inter-server Fairness Service // Proceedings of the 12th International Conference on Parallel and Distributed Systems (ICPADS'06). - 2006.
27. Chydzinski A. Nowe Modele Kolejkowe Dla Weztow Sieci Pakietowych. Gliwice: Pracownia Komputerowa Jacka Skalmierskiego, 2013
28. Floyd S., Jacobson V. Random early detection gateways for congestion avoidance. IEEE/ACM Transactions on Networking, 1993, vol. 1, №4, pp. 397–413.
29. Bohacek S., Shah K., Arce G. R., Davis M. Signal processing challenges in active queue management. IEEE Signal Processing Magazine Transactions on Networking, 2004, vol. 21, №5, pp. 69–79.
30. Bonald T., May M., Bolot J.-C. Analytic evaluation of RED performance. Proceedings of INFOCOM, 2000, vol. 3, pp. 1415–1424.
31. Tikhonenko O., Kempa W. M. The generalization of AQM algorithms for queueing system with bounded capacity. Lecture Notes in Computer Sciences, 2012, vol. 7204, pp. 242–251.
32. Tikhonenko O., Kempa W. M. Queue-size distribution in M/G/1-type system with bounded capacity and packet dropping. Communications in Computer and Information Science, 2013, vol. 356, pp. 177–186.

33. Kempa W. M. A direct approach to transient queue-size distribution in a finite-buffer queue with AQM. *Applied Mathematics and Information Sciences*, 2013, vol. 7, №3, pp. 909–915.
34. Боев В. Д. Моделирование систем. Инструментальные средства GPSS World. Санкт-Петербург: БХВ-Петербург, 2004.
35. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. М.: РУДН, 1995.
36. Гольдштейн Б.С., Соколов Н.А., Яновский Г.Г. Сети связи: Учебник для ВУЗов. - СПб.: БХВ - Петербург, 2010.
37. Жерновой Ю.В. Простейшие модели управления очередью в узлах сетей с коммутацией каналов//Информационные процессы, том 14, №2, 2014.
38. Вишнеvский В.М. Теоретические основы проектирования компьютерных сетей – М.: Техносфера, 2003.

Приложение А

Программа алгоритма RED в GPSS WORLD

Gau EQU 3 ; значение λ
Tal EQU 1 ; значение μ
TIME EQU 1000000 ; время моделирование
DISTR TABLE (Q\$QUE+S\$SYS) 0,1,1500 ; параметры распределения числа заявок
Prob VARIABLE N\$LB2/(2.3#N\$Lb0) ; вероятность обслуживания
SYS STORAGE 3 ; трехканальная система
GENERATE 1
TABULATE DISTR ; вычислить распределение числа заявок
TERMINATE
GENERATE (Exponential (5,0,(1/Gau))) ; поток заявок
Lb0 TEST E (S\$SYS+Q\$QUE),4,Lt5 ; проверка условия для задания β_4
TRANSFER 800,,Lbg ; задания β_4
TRANSFER, OUT
Lt5 TEST E (S\$SYS+Q\$QUE),5,Lt6 ; проверка условия для задания β_5
TRANSFER 400,,Lbg ; задания β_5
TRANSFER, OUT
Lt6 TEST GE (S\$SYS+Q\$QUE),6,Ltg ; проверка условия для задания β
TRANSFER 200,,Lbg ; задание β
TRANSFER, OUT
Lbg TRANSFER 200,,LQ ; задание α_1
TRANSFER 375,,L2 ; задание α_2
SPLIT 3,LQ ; задание α_3
TRANSFER, OUT
L2 SPLIT 2,LQ
TRANSFER, OUT
LQ QUEUE QUE
ENTER SYS
DEPART QUE
ADVANCE (Exponential(5,0,(1/Tal))) ; распределение времени обслуживания
Lb2 LEAVE SYS
OUT TERMINATE
GENERATE TIME
SAVEVALUE Prob, V\$Prob
TERMINATE 1
START 1

Приложение В

Алгоритм управления очередями в узле сети с коммутацией пакетов без повторного поступления в узел отброшенных пакетов в GPSS WORLD

```
MIN_TH EQU 60
MAX_TH EQU 100
DELAY TABLE M1,15,15,20
PROBAB VARIABLE (QA$OCHER-MIN_TH)/(MAX_TH-MIN_TH)
GENERATE (Exponential(1,0,0.7))
QUEUE OCHER
TEST G QA$OCHER,MIN_TH,OBSL
TEST L QA$OCHER,MAX_TH,SBROS
TEST LE V$PROBAB,0.95,SBROS
OBSL SEIZE UZEL
DEPART OCHER
ADVANCE (Exponential(1,0,1))
RELEASE UZEL
TABULATE DELAY
SBROS TERMINATE 1
START 1000
```

Приложение С

Алгоритм управления очередями в узле сети с коммутацией пакетов с повторным поступлением в узел отброшенных пакетов в GPSS WORLD

```
MIN_TH EQU 60
MAX_TH EQU 100
DELAY TABLE M1,15,15,20
PROBAB VARIABLE (QA$OCHER-MIN_TH)/(MAX_TH-MIN_TH)
GENERATE (Exponential(1,0,0.7))
GAUHAR QUEUE OCHER
TEST G QA$OCHER,MIN_TH,OBSL
TEST L QA$OCHER,MAX_TH,PEREM
TEST LE V$PROBAB,0.95,PEREM
OBSL SEIZE UZEL
DEPART OCHER
ADVANCE (Exponential(1,0,1))
RELEASE UZEL
TABULATE DELAY
TRANSFER ,SBROS
PEREM TRANSFER .300,SBROS,GAUHAR
SBROS TERMINATE 0
GENERATE 1000
TERMINATE 1
START 1
```