ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ

**Федеральное государственное
бюджетное образовательное учреждение высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА»**

О. А. Симонина

КАЧЕСТВО СЕРВИСОВ И УСЛУГ В СЕТЯХ СВЯЗИ

Учебное пособие

СПб ГУТ)))

**САНКТ-ПЕТЕРБУРГ
2016**

Рецензенты
главный научный сотрудник ЛО ЦНИИС
д. техн. наук, доц. Н.А. Соколов
доцент кафедры инфокоммуникационных систем СПбГУТ
к. техн. наук, В.С. Елагин

Рекомендовано к печати
редакционно-издательским советом СПбГУТ

Симонина, О. А.

Качество сервисов и услуг в сетях связи : учебное пособие / О. А. Симонина ; СПбГУТ. – СПб., 2016. – 80 с.

Написано в соответствии с программой дисциплины «Качество сервисов и услуг в сетях связи».

Предназначено для бакалавров четвертого курса по направлению подготовки очной и заочной формы обучения: 210700.62 «Инфокоммуникационные технологии и системы связи».

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Процессы изменения мира с точки зрения развития технологий, в том числе инфокоммуникационных, можно описать с использованием кондратьевских циклов. Позднее академиками Львовым Д.С. и Глазьевым С.Ю. было предложено использовать понятие технологического цикла, дополнившую теорию Н.Д. Кондратьева с точки зрения развития совокупности технологий и их жизненных циклов [1].

Рассмотрим технологические уклады XX – начала XXI века (рис. 1).

IV – первая половина XX века: развитие машиностроения, транспортной инфраструктуры, массового производства, энергетики.

V – вторая половина XX века: развитие телекоммуникаций, вычислительной техники, электроники.

VI – конец XX и первая половина XXI века: развитие нанотехнологий, новые подходы к природопользованию, биотехнологии.

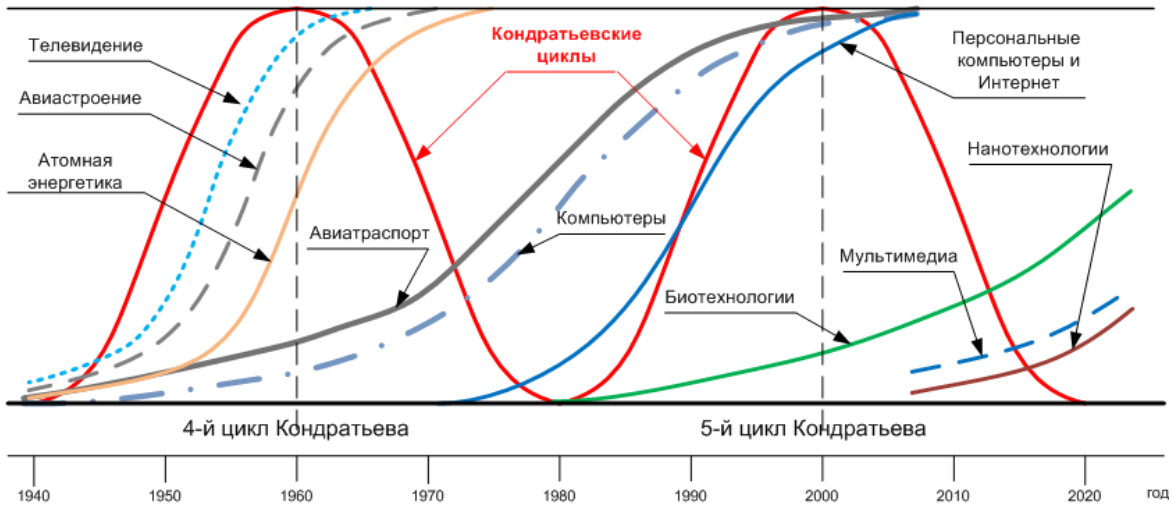


Рис. 1 - Инфратраектории некоторых макротехнологий современной экономики [2]

Телекоммуникации являются одной из относительно молодых технологий, цель которой заключалась в создании технических средств передачи информации от отправителя к получателю в целости и в максимально короткие сроки. С момента изобретения радио и телефона и до начала XXI века от средств связи требовалось не более этого. Однако уже в 70-е годы XX века появилась идея отойти от традиционного подхода, когда каждая сеть отвечала только за одну услугу, и начать разработку мультисервисных сетей. Именно тогда появились технологии ISDN и, чуть позже, ATM. Идеи, которые родились при разработке этих технологий, успешно применяются и в современных пакетных сетях. Если трафик традиционных сетей разделялся на пользовательский и трафик сигнализации, причем для сигнализации строилась своя сеть, то уже в ATM пользовательский трафик представлен несколькими типами: реального времени с постоянной скоростью, реального времени с переменной скоростью, близкого к реальному времени, передачи данных. Это разделение позднее, с приходом IP-сетей, постепенно трансформировалось в концепцию Triple Play: реальное время, поток, передача данных.

Дальнейшее развитие в направлении NGN привело к разделению трафика передачи данных на два типа: чувствительного и нечувствительного к задержкам. Часть приложений порождала трафик, вполне соответствующий традиционному взгляду на передачу данных – гарантированная доставка, перезапросы и, как следствие, низкая чувствительность к задержкам при недопустимости потерь. Однако стали появляться приложения, требующие низких значений задержек, такие как процессы управления on-line. Одновременно появляются устройства, генерирующие трафик для создания комфортной среды, непосредственно не затрагивающий органы чувств человека. Такой трафик получил название mashine-to-mashine или M2M.

Именно в этот момент, на стыке развития телекоммуникаций и других отраслей, особенно робототехники и биологии, когда рынок телекоммуникаций стал насыщен мультисервисными услугами, появляется новая концепция. Теперь речь идет о создании единого инфокоммуникационного пространства, развитии мультимедиа и виртуальной реальности, Интернета вещей и автоматизированной среды.

Итак, рассмотрим эволюцию технологий телекоммуникаций:

1. Традиционные сети: качество услуги полностью определяется особенностями технологии (ТфОП, FR, 2G и т.д.)
2. Мультисервисные сети (Triple Play): передача по одной сети нескольких типов трафика (ISDN, ATM, 3G).
3. Сети NGN: управление мультисервисными услугами не зависимо от технологии (IMS, SIP). Мобильность пользователя.
4. Сети NGN2 или post-NGN: приоритет беспроводного доступа, создание единого инфокоммуникационного пространства, взаимопроникновение идей и технологий автоматизации и телекоммуникаций, наносети, самоорганизация, поиск новых механизмов и сред передачи информации.

В настоящее время решены задачи, стоящие перед NGN:

- Независимое развитие уровней согласно модели NGN: отделение услуг от сетевой технологии.
- Мобильность пользователя.
- Мультисервисность.
- Сращивание телекоммуникационных и компьютерных технологий.
- Развитие технологий виртуализации.

Можно сказать, что мир входит в новую эпоху – сетей после NGN: post-NGN или NGN2.

При классификации современных сетей связи имеет смысл обратить внимание на разделение по типу коммутации (рис. 2). На начальном этапе различались только крайние случаи – коммутация пакетов и коммутация каналов. Позднее, с развитием сетей передачи данных, появилось такое понятие как коммутация сообщений. На несколько десятилетий этот вид коммутации был забыт, но сейчас, с приходом Интернета вещей, интерес к нему снова возвращается. Важно отметить, что существует еще гибридный вид коммутации: коммутация пакетов в поддержкой виртуальных каналов. По такому принципу функционируют практически все современные мультисервисные технологии.

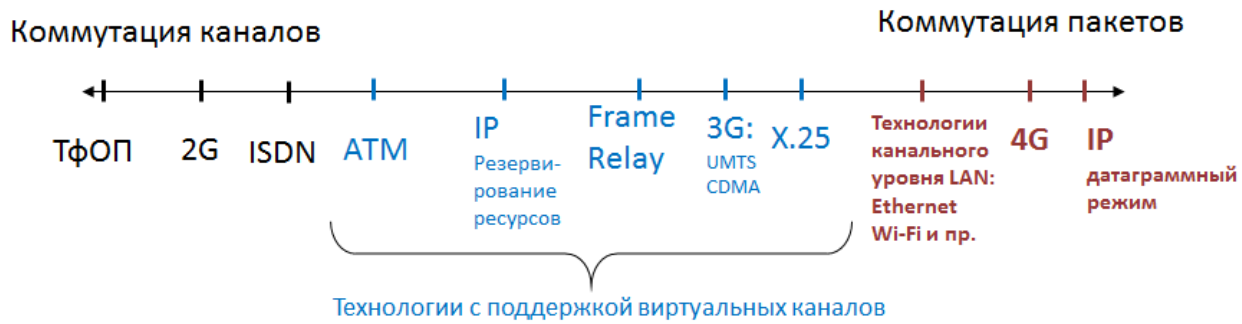


Рис. 2 – Классификация технологий телекоммуникаций по типу коммутации

Для описания архитектуры сетей используются базовые топологии (*шина, звезда, кольцо*), типы рассылки (*unicast*, *multicast*, *anycast*), протоколы построения деревьев, принципы иерархического построения сети.

Базовое иерархическое построение сети предполагает наличие максимум трех уровней иерархии: ядро сети, уровень агрегации и уровень доступа (рисунок 3). Отметим, что в небольших сетях может происходить вырождение иерархии до глубины в два уровня: доступ и ядро.

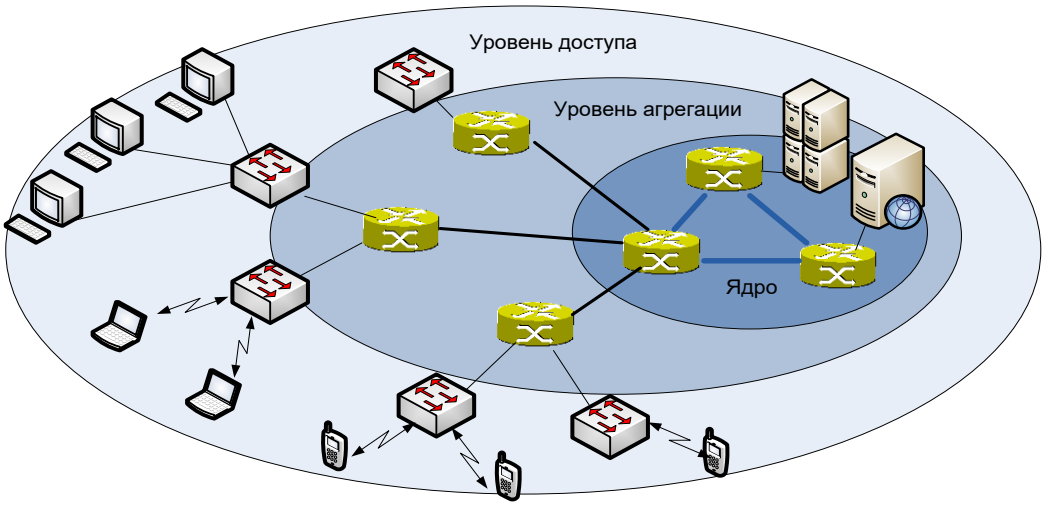


Рис. 3 – Базовая иерархическая модель сети

В ядре, как минимум, находятся опорные маршрутизаторы, могут быть расположены граничные маршрутизаторы и серверы услуг. Уровень агрегации обеспечивает возможность агрегации и распределения трафика внутри сети оператора, может выполнять роль интегратора отдельных сегментов сети. Уровень доступа организует возможность физического доступа пользователя к сети оператора.

Подобный подход не зависит от используемой оператором технологии, но позволяет эффективно выстроить архитектуру, повысив устойчивость сети, и сделать процессы управлении сетью прозрачными.

Таким образом, все сетевые элементы современных сетей связи могут быть классифицированы по принадлежности к иерархическим уровням. В зависимости от уровневой принадлежности к сетевым элементам предъявляются различные требования по набору функциональных возможностей, надежности, производительности и др. Традиционно оборудование разделяют на оборудование операторского класса и клиентское. Клиентское находится на стороне пользователя, позволяя ему организовать доступ в сеть, управляется непосредственно пользователем и/или оператором. Оборудование операторского класса формирует сеть связи оператора и управляется только оператором.

Базовой сетевой моделью традиционно считают семиуровневую эталонную модель взаимодействия открытых систем ISO/OSI, разработанную в 1978 году на базе протокола *X.25*. Данная модель является удобным абстрактным инструментом при разработке сетевых протоколов и представлении процессов в сетях связи не зависимо от технологий, размеров и конфигурации сетей.

С появлением IP-сетей была разработана четырехуровневая модель TCP/IP. Особенностью модели было объединение физического и канального уровней в уровень подсетей, а трех верхних уровней модели ISO/OSI – в единый прикладной уровень. Данное представление было призвано отразить независимость прокола IP как от технологии передачи, так и от вида услуги. Однако для эксплуатации данное представление оказалось недостаточным и привело к появлению, так называемой, гибридной модели TCP/IP, пятиуровневой, с выделенными канальным и физическим уровнем.

С переходом к сетям NGN появилась необходимость развития концепции независимого и «бесшовного» взаимодействия уровней, при котором учитывалось отдельное от инфраструктуры операторов развитие услуг и приложений. Это породило трехуровневое представление: уровень услуг, уровень управления соединениями, транспортный уровень. При этом последний включает в себя всю инфраструктуру сети и не зависит от вида предоставляемых услуг и управления ими. На рисунке 4 представлено соответствие рассмотренных моделей.

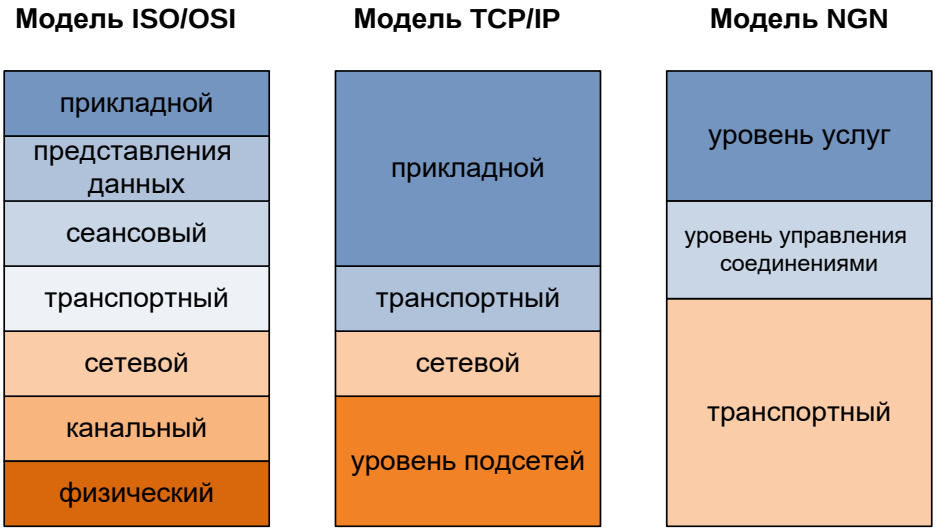


Рис. 4 – Соответствие эталонной модели *ISO/OSI*, модели *TCP/IP* и модели *NGN* по уровням [3]

В зависимости от ситуации и технологической необходимости в настоящее время используется каждая из этих моделей. При этом легко заметить, что модель NGN отражает, в первую очередь, взаимодействие операторов сетей и операторов услуг, а также процессы на сети оператора, предоставляющего услуги высоких уровней (кроме доступа в сеть). Модель TCP/IP хорошо описывает процессы в сетях, ориентированных на логические соединения. Эталонная модель ISO/OSI ориентирована на технические решения и незаменима в процессе эксплуатации, при разработке систем управления сетью и услугами и пр., несмотря на то, что некоторые современные решения не полностью соответствуют первоначальным определениям функционала уровней.

Отметим, что данные модели не дают представления об архитектуре сети, а отражают только принципы взаимодействия протоколов и последовательность обработки информации на узлах.

1 УСЛУГИ В ИНФОКОММУНИКАЦИЯХ

1.1 Стандартизация в области качества

Функционирование сетей связи невозможно без единых стандартов, определяющих архитектуру, протоколы, технологии, требования. Кроме национальных стандартов, действующих в пределах одного региона, необходимо развитие международных стандартов и рекомендаций, позволяющих организовывать взаимодействие сетей по всему миру. Международные организации, стандартизирующие решения в области телекоммуникаций:

- **ITU-T:** International Telecommunication Union – Международный Союз Электросвязи (МСЭ)**.** Рекомендации имеют обозначения «буква.число» (например, Y.1540).
- **IETF:** Internet Engineering Task Force – Инженерная группа по решению задач Internet. Рекомендации носят формат «RFC число» (например, RFC 821)
- **ETSI:** European Telecommunications Standardizations Institute - Европейский институт по стандартизации телекоммуникаций

Также многие технологии описаны в рекомендациях IEEE – Institute of Electrical and Electronics Engineers. Стандарты носят формат «IEEE.число» (например, IEEE.802.3). IEEE входит в **NIST:** National Institute of Standarts and Technology – Национальный институт стандартов США.

В Российской Федерации основными документами являются Закон о связи, Приказы Минкомсвязи и ГОСТы. В связи с глобальностью телекоммуникаций большинство ГОСТов представляют собой или версию международного стандарта, или адаптацию международного стандарта к реалиям страны.

Для всех участников телекоммуникационного рынка базовыми рекомендациям в области качества являются рекомендации МСЭ:

- **E.800-E.899**: Quality of telecommunication services: concepts, models, objectives and dependability planning
- **Y.1500-Y.1599**: Quality of service and network performance
- **Y.2100-Y.2199**: NGN: Quality of Service and performance

- **G.1000-G.1999**: Multimedia Quality of Service and performance – Generic and user-related aspects
- **E.490-E.799**: Traffic engineering

Подробнее с ними можно ознакомиться на сайте МСЭ в разделе «Рекомендации» [16].

1.2 Подходы к оценке качества услуг

С точки зрения предоставления услуг имеет смысл говорить об инфокоммуникационных услугах, ориентированных не только на передачу информации. Вопросы хранения контента, доступа к контенту, предоставления услуг из нетрадиционных для операторов связи областей являются приметами конца XX-начала XXI века.

Основные определения

Предоставление услуг - любая деятельность, связанная с предоставлением услуги поставщиком услуг, с момента поступления заказа на услугу до момента наличия услуги для ее использования абонентом/пользователем.

Услуги - результаты непосредственного взаимодействия поставщика и потребителя и внутренней деятельности поставщика по удовлетворению потребностей пользователя.

Сервисы – совокупность услуг и правил их использования.

Мультисервисные услуги - услуги электросвязи, которые поддерживают одновременное использование многих типов средств информации (речь, данные, видео).

Обслуживание – набор функций, предоставляемых пользователю организацией.

Пользовательская инфокоммуникационная среда - совокупность средств реализации инфокоммуникационных технологий (аппаратных и программных средств вычислительной техники и связи), услуг и информационных ресурсов.

Инфокоммуникационная структура сети электросвязи - совокупность информационных ресурсов и инфраструктуры сети электросвязи.

Инфокоммуникационная услуга (услуга информационного общества) - услуга связи, предполагающая автоматизированную обработку, хранение или предоставление по запросу информации с использованием средств вычислительной техники, как на входящем, так и на исходящем конце соединения

Сервисные платформы - совокупность технологий и программных решений, позволяющая осуществлять разработку и/или использование программных продуктов с использованием модульного подхода.

Сервисная информация - Информация о службе, позволяющая принимать данные этой службы

В мультисервисных сетях традиционные и новые услуги могут быть созданы на базе одной сетевой технологии (IP), использовать различные пользовательские платформы. Телекоммуникационные операторы, средства массовой информации и IT-компании могут использовать новые технологии и новое регулирование для развития услуг вне своих традиционных бизнес-секторов.

Задача обеспечения качества услуг: передать трафик с параметрами, позволяющими восстановить информацию на приеме с минимально заметными для пользователя изменениями. Она может быть разбита на несколько этапов:

1. Определить параметры, по которым может быть оценено качество услуги.
2. Определить нормы, определяющие незаметность изменений для пользователя.
3. Обеспечить соблюдение норм при передаче на сети (разработка механизмов обеспечения качества).
4. Обеспечить меры по нивелированию влияния сети на приеме.

Концепции оценки качества услуг

На основе этого подхода была создана концепция Triple Play, классифицирующая трафик по трем базовым типам: реального времени (телефония), потоковый (видео), эластичный (трафик передачи данных: нормального приоритета и фоновый). Различные вариации этого подхода и классификация трафика в их системе оценок будут подробно рассмотрены далее.

Отметим, что QoS не является единственной системой оценок, позволяющей оценить качество современных инфокоммуникационных услуг. Например, предлагается сквозная модель [4], учитывающая и качество восприятия, и бизнес-ориентированные показатели (рис. 5).

Рис. 5 – Структура обеспечения инфокоммуникационных услуг

С 2009 года существует концепция Интернета вещей, объединяющая классический подход Triple Play и элементы автоматизации на новом эволюционном витке. Оценка качества услуг в рамках новой парадигмы требует учитывать и межмашинное взаимодействие, и создание вокруг человека новой среды, не только инфокоммуникационной, то и физической, контролируемой техническими средствами.

Таким образом, можно сформулировать следующие проблемы в области обеспечения качества современных инфокоммуникационных услуг:

- Появление новых услуг и сервисов, для которых не определены нормы показателей качества обслуживания.
- Широкое распространение беспроводных технологий, имеющих слабую поддержку QoS.
- Увеличение мультимедийной составляющей трафика, особенно видео HD.
- Распространение услуг OTT и smart-устройств.
- Неравномерность развития инфраструктуры операторов связи.
- Компьютеризация телекоммуникационного оборудования и возрастающее влияние информационной безопасности на качество телекоммуникационных услуг.

1.3 Показатели QoS

Показатели качества обслуживания (QoS – Quality of Service) являются важными критериями оценки качества и описаны в Рекомендации Y.1540:

- Задержка в сквозном соединении и джиттер* задержки (мс).
- Величина потерь (%).
- Производительность сети (бит/с).
- Надёжность сетевых элементов (Кг).
- Устойчивость (живучесть) сети - возможность сохранения функционала при выходе из строя отдельных элементов

Согласно рекомендации Y.1541 весь трафик можно разделить на классы:

Класс 0: Приложения реального времени, чувствительные к джиттеру, характеризуемые 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 приложения

Класс 4: Приложения, допускающие низкий уровень потерь (короткие транзакции, массивы данных, потоковое видео)

Класс 5: Традиционные применения сетей IP

Эта же рекомендация задает нормы на параметры доставки IP-пакетов с разделением по классам обслуживания (табл. 1)

Таблица 1 – Классы QoS согласно Y.1541 (Н – не нормировано)

Сетевые характеристики	**Классы QoS**					
	0	**1**	**2**	**3**	**4**	**5**
Задержка доставки пакета IP, мс	100	400	100	400	1000	Н*
Вариация задержки пакета IP, мс	50	50	Н	Н	Н	Н
Коэффициент потери пакетов, 10^{-3}	1	1	1	1	1	Н
Коэффициент ошибок пакетов, 10^{-4}	1	1	1	1	1	Н

2 КАЧЕСТВО ВОСПРИЯТИЯ МУЛЬТИМЕДИЙНЫХ УСЛУГ В NGN

Термин «качество восприятия» - QoE, Quality of Experience - был введен специально для оценки видео. Однако, он может быть применен и к другим типам мультимедийных услуг, ориентированным на восприятие человеком, таких как аудио и виртуальная реальность. Работы по созданию виртуальной реальности пока находятся в состоянии, не имеющем научной базы для проведения таких исследований. Передача аудио поверх IP имеет длительную историю роста и становления, к 1995 году IP-телефония вошла в стадию зрелости. Первая часть лекции посвящена вопросам формирования передачи аудио поверх IP-сетей, методам оценки качества и проблемам нивелирования потерь.

2.1 Audio over IP

Формирование речи

Восприятие человеком звуков не зависит от способа их передачи. Поэтому основной идеей при передаче аудио, как речи, так и музыки, является организация синхронного потока поверх асинхронной IP-сети. В современных сетях аудио-сервисы можно разделить на следующие два типа:

1. Телефония - voice over IP (VoIP):
   - Частота дискретизации 8 кГц.
   - Поддержка реального времени (см. нормы).
   - Организация дуплекса.
   - Отсутствие перезапросов для потерянных пакетов.
2. Вещание:
   - Частота дискретизации до 44 кГц.
   - Условно реальное время (значения задержек до 500 мс).
   - Симплекс.
   - Перезапросы.

Для начала рассмотрим основные численные параметры, характеризующие воспринимаемый аудиодиапазон (рис. 6):

- Воспринимаемый звуковой диапазон 20 Гц-20 кГц
- Диапазон максимальной чувствительности слуха 2-5 кГц
- Ухо человека не чувствительно к фазовым искажениям.
- Диапазон речевого сигнала 35-7000 Гц
- Частота основного тона 70-450 Гц: мужчины 100-200 Гц (чаще всего); женщины 150-300 Гц (чаще всего); дети 300-450 Гц.

Все звуки, которые издает человек, можно разделить на:

- Гласные: всегда огласованные - в формировании звука участвуют голосовые связки и органы речевой полости (зубы, губы, язык), высокая длительность звука.
- Согласные (сонорные, щелевые, взрывные, аффрикаты):
  - огласованные: возбуждение сигналом основного тона - в формировании звука участвуют голосовые связки и органы речевой полости (зубы, губы, язык);
  - неогласованные: возбуждение шумовым сигналом - в формировании звука участвуют только органы речевой полости (зубы, губы, язык).

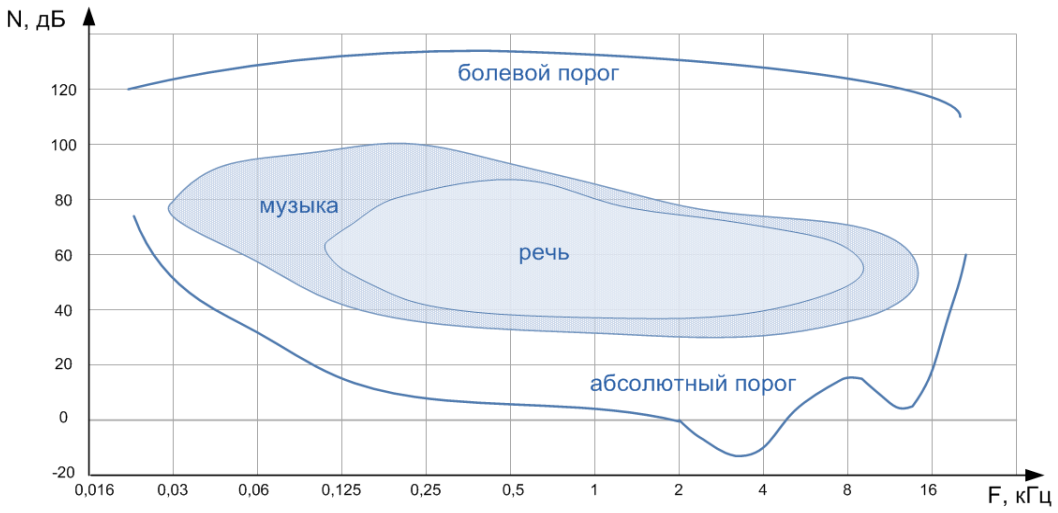


Рис. 6 – Область слышимых звуков [5]

Если представить речевой тракт в виде резонансной системы с двумя источниками возбуждения: сигналом основного тона или шумовым сигналом (рис. 7). Источником шума являются легкие – воздух выходит из них и по пути встречает различные препятствия в виде органов ротовой полости. Если при этом задействованы голосовые связки, то их колебания можно рассматривать как частоту основного тона. Для огласованных звуков можно во временной области выделить период основного тона: $T_{от}=1/F_{от}$ (рис. 8). Частота основного тона всегда находится в динамике. При этом в качестве резонаторов выступает грудная клетка для шумовых сигналов и черепная коробка для сигналов с возбуждением основным тоном, F_n – частоты формант.

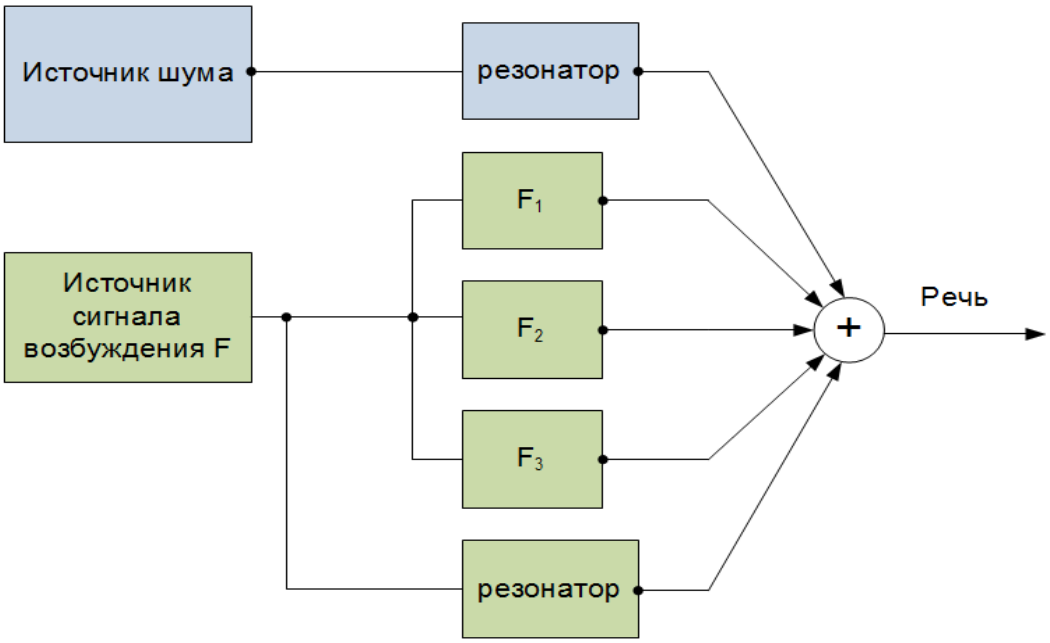


Рис. 7 – Модель формирования речевого сигнала

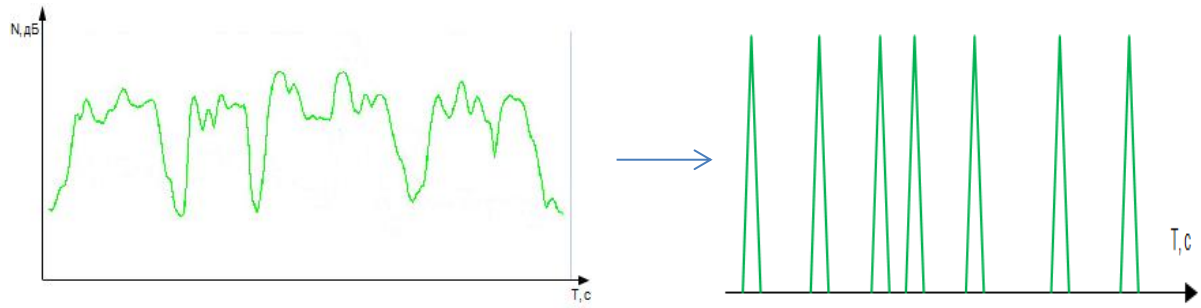


Рис. 8 – Выделение частоты основного тона

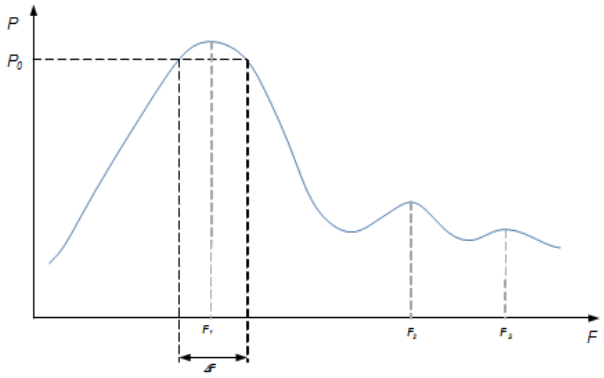


Рис. 9 – Форманты речевого сигнала

Одним из основных методов анализа характеристик речевого сигнала является форматный анализ. Форманта – это область наибольшей энергии сигнала. Для речевых сигналов характерно наличие трех формант (рис. 9), при этом у гласных энергия в основном сосредоточена в области первой форманты, а для шипящих – в области третьей, хотя в целом можно сказать, что шипящие согласные имеют сглаженный спектр без ярко выраженных формант.

Передача речи по пакетной сети связи

Рассмотрим организацию процесса передачи речи по пакетной асинхронной сети (рис. 10). На стороне отправителя находятся функциональные блоки, обеспечивающие функции анализатора, на стороне получателя – синтезатора.

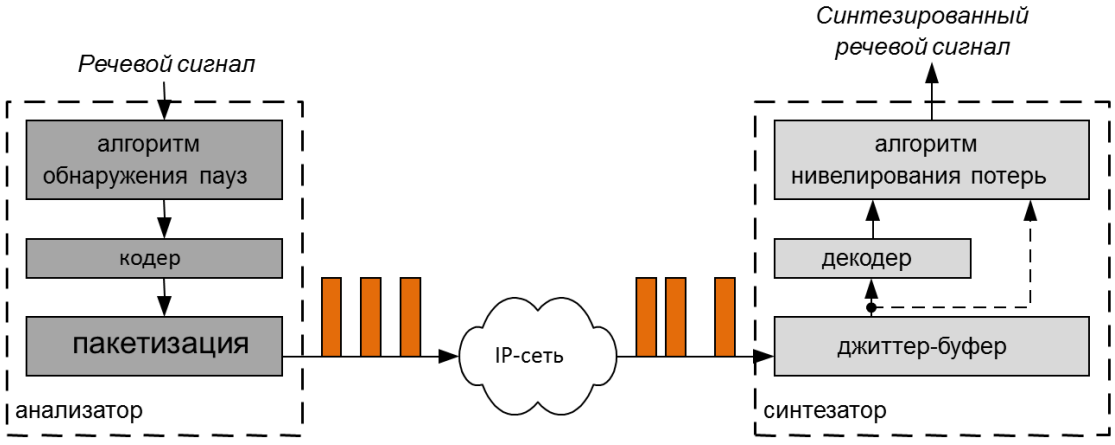


Рис. 10 - Организация передачи речи по IP-сети

Речевой сигнал поступает на обработку и может быть подвергнут выделению пауз, так как речь приблизительно на 60% состоит из периодов молчания (вдох, прослушивание собеседника, размышление). Алгоритм выделения пауз важен для подавления периодов молчания (рис. 11). Для экономии пропускной

способности используется набор алгоритмов, позволяющих не передавать «тишину», а использовать освободившийся ресурс для передачи данных:

- VAD (Voice Activity Detector) – обнаруживает период молчания.
- DTX (Discontinuous Transmission) – прекращает передачу в период молчания или генерирует специальные пакеты-идентификаторы паузы и «тишина». Во втором случае эффективность использования пропускной способности ниже.
- CNG (Comfort Noise Generator) – заменяет на стороне приема тишину на окрашенный шум, близкий к фактическому.

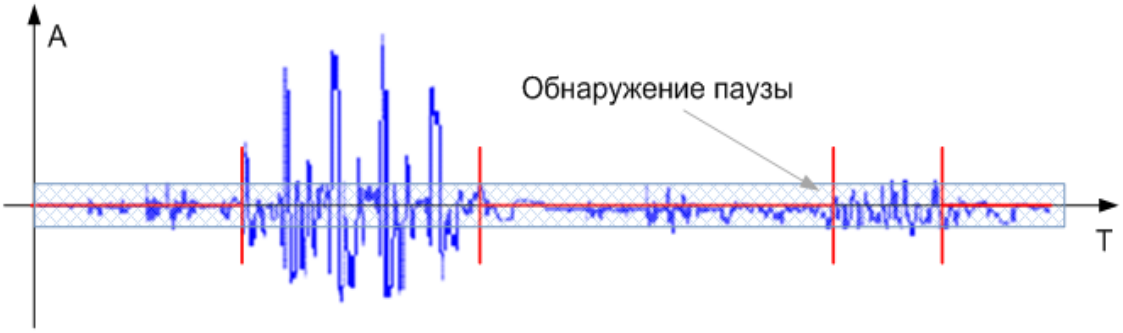


Рис. 11 – Подавление периодов молчания

Отметим, что использование этих алгоритмов ухудшает качество речи, вплоть до 30%. Причины ухудшения качества:

- не подключен генератор комфортного шума — создается эффект «гробовой тишины», особенно если собеседник находился в зашумленной обстановке;
- ложное срабатывание при обнаружении тишины.

Использование этих алгоритмов не является обязательным, поэтому для обеспечения более высокого качества речи можно избежать их применения.

Далее сигнал подвергается кодированию. Скорость передачи современных речевых кодеков 1,2-64 кбит/с. Рассмотрим основные типы кодеков, применяемых в IP-телефонии, и их характеристики качества.

В частотной области, на основе импульсно-кодовой модуляции популярны стандарты:

- G.711 – ИКМ;
- G.722, G.726 – АДИКМ.

В спектральной области используются стандарты: G.723, G.728, G.729, iLBC, GSM и пр., в том числе проприетарные. В целом их можно разделить на следующие типы:

- вокодеры (полосовые кодеки);
- LPC (на основе линейного предсказания);
- CELP (на основе линейного предсказания с возбуждением от кодовой книги).

Кодеки семейства ИКМ работают в частной области. Для кодирования производятся процедуры дискретизации и квантования, имеют ограничение согласно теореме Котельникова: при частоте дискретизации *Fd*=8кГц рабочий частотный диапазон 0-4 кГц, откуда диапазон канала тональной частоты (телефонного) 0,3-3,4 кГц. Как следствие, форманты некоторых шипящих согласных лежат за пределами рабочего диапазона, что ухудшает разборчивость речи.

Кодек G.711 использует импульсно-кодовую модуляцию, скорость кодека 64кбит/с, что соответствует каналу E0. Этот кодек поддерживается всеми системами телефонии в обязательном порядке в целях совместимости решений различных производителей.

Кодек G.722 используется в аудиосистемах, на радиостанциях, для трансляции речи, совместим с потоковым аудио. Семейство кодеков включает:

- G.722, 1988 г. Технология АД-ИКМ, скорости 48, 56 и 64 кбит/с;
- G.722.1, 1999 г., скорости 24 и 32 кбит/с, *Fd*=16кГц;
- Annex C — 48 кбит/с, *Fd*=28кГц;
- G.722.2, 2002 г. Технология AMR-WB (Adaptive Multi Rate — Wide Band) это разработка 3GPP; скорости от 6,6 кбит/с до 23,85 кбит/с, *Fd*=28кГц.

Особенности G.722: сопряжение с ИКМ, 8 кГц, A- и μ-закон, эхокомпенсация согласно рекомендации G.168.

В аналоговом сигнале, содержащем речевую информацию, невозможны резкие скачки интенсивности, поэтому можно кодировать изменения сигнала относительно предыдущего значения, а не мгновенное значение амплитуды, что легло в основу технологии АД-ИКМ. Таким образом, можно использовать меньшее количество бит (от 4-х) для кодирования, что дает существенное понижение скорости кодека. Задержка квантования при этом сохраняется несущественной (0,125 мкс).

Широкое распространение получили кодеки на основе линейного предсказания (LPC – Linear Coder Prediction). К их достоинствам относится хорошая компрессия речи, позволяющая существенно понизить скорость кодека и, как следствие, требования к пропускной способности сети. Созданию таких кодеков способствовала особенность слуха человека – наше ухо не чувствительно к фазовым искажениям. Архитектура кодеков на основе линейного предсказания приведена на рис. 12.

Рассмотрим принцип кодирования в кодеках LPC. Метод линейного предсказания базируется на представлении значения стохастического сигнала *S(t)* линейной суммой его предыдущих значений, умноженных на некоторые коэффициенты *a(i)*:

$$S(t) = -\sum_{i=1}^{p} S(t-i) \cdot a(i) + Gu_n,$$

где *G* - коэффициент усиления.

Применив z-преобразование, можно получить передаточную функцию *H(z)* следующего вида:

$$H(z) = \frac{1}{A(z)} = \frac{G}{\sum_{i=1}^{N} a_i z^i}.$$

Функцию *A(z)* называют обратным фильтром, или фильтром предсказателем (рис.12).

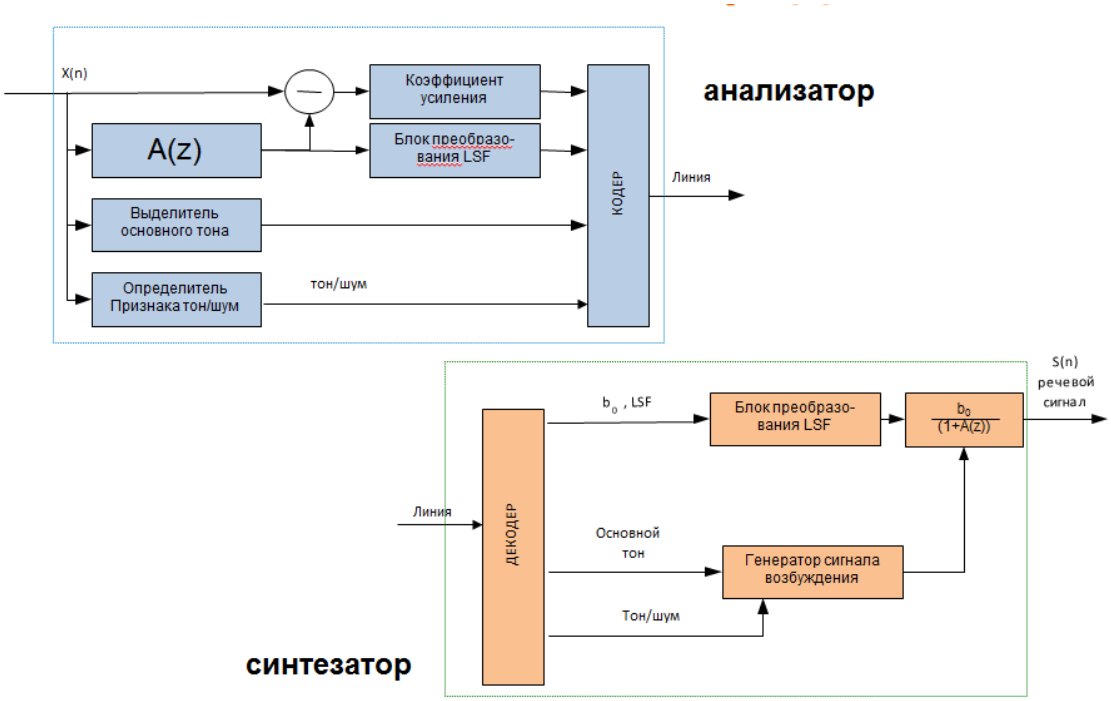


Рис. 12 - Архитектура кодека на основе линейного предсказания

Таким образом, зная параметры фильтра a_i и *G*, а также то, что сигнал *S(t)* является стационарным на некотором интервале времени, можно сгенерировать данный сигнал по его начальным значениям.

Модель линейного предсказания, проводимая во временной области, может быть проведена с применением преобразований Фурье также и в частотной области. Это означает, что метод линейного предсказания сводится к корреляционному анализу, который может быть выполнен как во временной, так и в частотной области. На практике для расчета параметров речевого сигнала широкое применение получил метод Итакуры, заключающийся в том, что из полинома *A(z)* получают полиномы *P(z)* и *Q(z)*, такие, что:

$$A(z) = \frac{P(z) + Q(z)}{2}.$$

Корни полиномов *P(z)* и *Q(z)* лежат на единичной z-плоскости и могут быть записаны через полярные координаты: $z_P = e^{jw_{P_i}}$ и $z_Q = e^{jw_{Q_i}}$, где w_P и w_Q - линейные спектральные корни:

$$0 < w_{P_1} < w_{Q_1} < w_{P_2} < w_{Q_2} < \ldots < w_{P_p} < w_{Q_p} < \pi .$$

Частоты w_P и w_Q математически эквивалентны коэффициентам линейного предсказания a_i, но более устойчивы к квантованию. Эти частоты получили название линейные спектральные частоты (Linear Spectrum Frequency - LSF) или, в отечественной литературе, линейные спектральные корни. На приеме по ним эффективно восстанавливается полином *A(z)*. Особенностью таких корней являются их низкая чувствительность к квантованию, чередуемость и локальное влияние на спектр восстанавливаемого сигнала.

При использовании кодовой книги (CELP) 10 линейных спектральных корней заменяются на номер в строке, соответствующий набору близких по значению LSF (рис. 13). Для построения кодовой книги применяется метод векторного квантования. Конечно, чем меньше размер кодовой книги, тем ниже скорость кодека и тем хуже качество сигнала (по узнаваемости прежде всего).

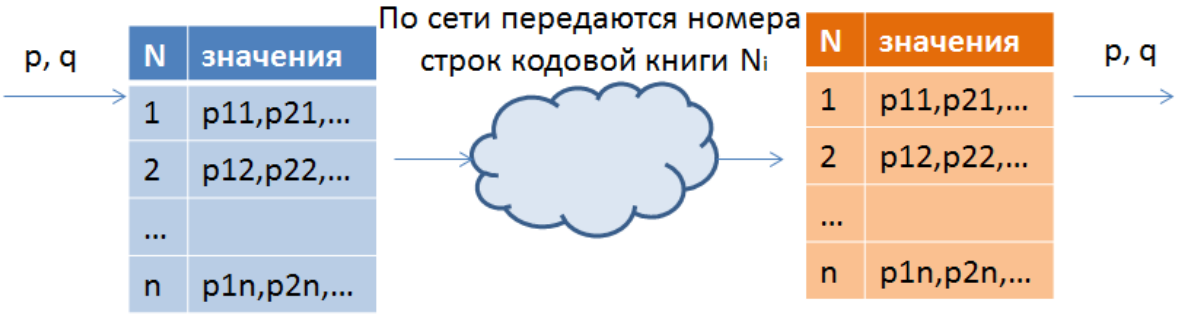


Рис. 13 – Принцип формирования CELP

На основании параметров линейного предсказания LSF, коэффициента усиления *G*, частоты основного тона $f_{от}$ и признака тон/шум формируется пакет данных на выходе кодека.

В зависимости от типа кодека пакет может иметь различную длину (рис. 14) за счет различного подхода к кодированию огласованных и неогласованных участков.

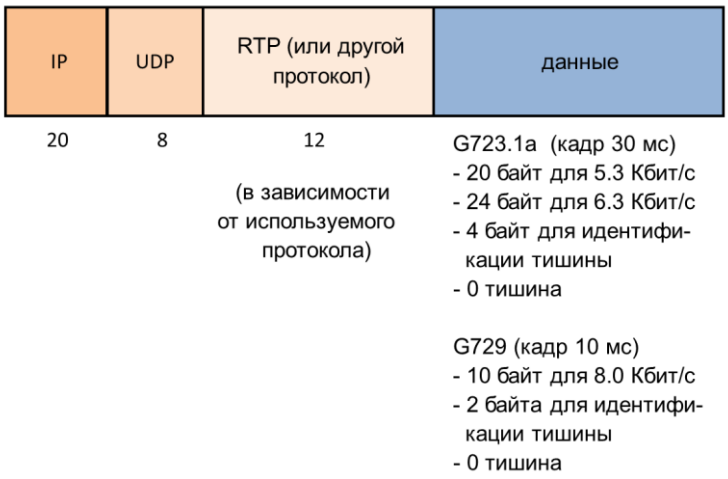


Рис.14 - Структура речевого пакета на выходе LCP-кодека

Речевые фрагменты в зависимости от участка речи можно разделить на кадры, содержащие речь, и кадры, содержащие «тишину», т.е. в которые попадают паузы. В среднем на каждые 352 мс речи приходится 650 мс пауз. Учитывая такую диспропорцию речи, для передачи пауз используются короткие пакеты, не содержащие данных в соответствующем поле, или содержащие 2 или 4 бита (в зависимости от типа кодека) для определения начала паузы. Для удобства пользователя на приемной стороне на данных отрезках будет генерироваться комфортный шум. Следовательно, речевой IP-пакет может быть различной длины в зависимости от компрессируемого участка речи. Постоянная составляющая длины пакета – это служебная информация, например: заголовок IP (20 бит), заголовок UDP (8 бит) и заголовок RTP (12 бит). Вместо RTP часто может использоваться другой протокол, реализованный в данной системе в зависимости от фирмы-производителя. В таком случае количество бит служебной информации может быть еще меньше в зависимости от заголовка данного протокола реального времени. Для уменьшения полосы пропускания, занимаемой одним речевым потоком, также может использоваться протокол сжатия заголовков RTP-пакетов CRTP (Compressed Real-time Transport Protocol), позволяющий уменьшить суммарный заголовок IP/UDP/RTP до 2—4 байт.

Таким образом, анализатор кодека формирует IP-пакеты переменной длины на основании постоянных по длительности кадров речи, обеспечивая постоянную задержку анализатора кодека t_{an}, прямопропорционально зависящую от размера кадра кодека:

$$t_{an} = t_{frame} + t_{al},$$

где t_{fr} – размер кадра, t_{al} – задержка алгоритма кодека или пакетизации.

Оценка качества речи

Раннее уже рассматривались нормы качества обслуживания для различных классов трафика согласно Рекомендации Y.1541. Однако для IP-телефонии на практике часто используют нормы ETSI, предполагающие более тонкое разделение (табл. 2 и 3).

Таблица 2 - Нормы QoS для IP-телефонии согласно ITU-T

Сетевые характеристики	**Классы QoS (real time)**	
	0 – высокий уровень интерактивности	**1 - интерактивные**
Задержка доставки пакета IP, мс	100	400
Вариация задержки пакета IP, мс	50	50
Коэффициент потери пакетов, 10^{-3}	1	1
Коэффициент ошибок пакетов, 10^{-4}	1	1

Таблица 3 - Нормы QoS для IP-телефонии согласно ETSI

Показатели качества	**Класс 1 - gold**	**Класс 2 - silver**	**Класс 3 - bronze**
Задержка, мс	< 150	150-250	250-400
Джиттер задержки, мс	< 10	10-20	20-40
Потери, %	< 0.5	0.5 - 1	1 - 2

Для оценки качества аудио при передаче по сети используют два типа методов:

1. Субъективные методики оценки качества:

– артикуляционные методы;

– MOS - Mean Opinion Score – восприятие качества услуги пользователем по 5-бальной шкале. Не учитывает особенности передачи речи по сети.

Для субъективной оценки согласно MOS обычно используют национальные стандарты, так как должны быть учтены языковые особенности. Для тестирования разрабатываются стандартные фразы. Рекомендуемое количество экспертов: 10-12 человек, обязательно мужчины и женщины, количество дикторов не менее 5. Экспертам должен быть знаком голос диктора. Для оценки используется специальные бланки по форме, согласно Р.48. Оценка производится или по критерию лучше/хуже или в баллах по трем типам: по узнаваемости, по разборчивости и общая.

Основой артикуляционных оценок качества речи является отношение правильно принятых элементов речи к их общему числу, вычисленному на репрезентативной выборке. Наиболее часто в качестве элементов рассматриваются слоги и слова. Разборчивость слов оценивается по итогам диагностических испытаний (DRI): максимальная разборчивость 80%, максимальная слоговая разборчивость в кодеках 87%. При этом данная методика не учитывает общее впечатление от речи: тембр, натуральность и т. д.

2. Объективные методики оценки качества:

Е-модель – многокритериальная оценка качества (табл. П1) в диапазоне от 1 до 100 баллов, описана в рекомендациях G.107, G.109.

$$R = R_o - l_s - l_d - l_e + A$$

где R_o = 93,2 – базовое значение R-фактора (по результатам работы кодека), l - искажения, вносимые сетью. Коэффициенты l_i обозначают искажения, вносимые трактом передачи.

Выведена зависимость между MOS и R-фактором (рис. 15), позволяющая определить по полученному значению Е-модели приблизительную оценку MOS (табл. 4).

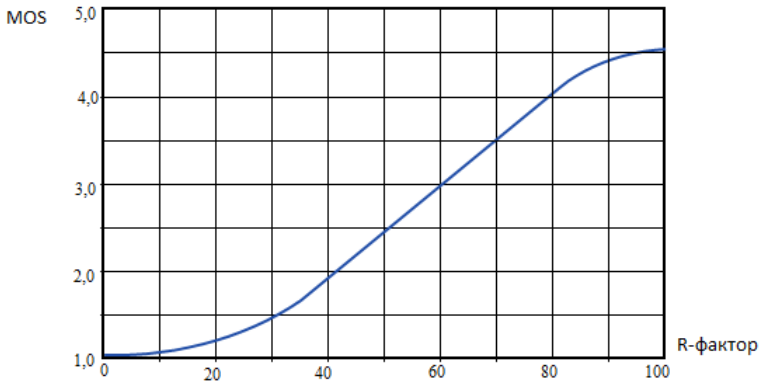


Рис. 15 – Соотношение R-фактора и MOS

Таблица 4 - Соотношение R-фактора и MOS

R-фактор	**Категория качества и оценка пользователя**	**MOS**
90<R<100	Самая высокая	4,34 – 4,50
80<R<90	Высокая	4,03 – 4,34
70<R<80	Средняя (часть пользователей оценивает качество как неудовлетворительное)	3,60 – 4,03
60<R<70	Низкая (большинство пользователей оценивает качество как неудовлетворительное)	3,10 – 3,60
50<R<60	Плохая (не рекомендуется)	2,58 – 3,10

Все стандартизированные кодеки имеют определенные значения MOS, учитывающие эффекты кодирования и компрессии речи (примеры приведены в табл. 5). Заметим, что максимальная задержка джиттер-буфера для всех типов кодеков составляет удвоенную длительность кадра.

Таблица 5 – Параметры качества некоторых стандартизированных кодеков

Кодек	**G.711**	**G.723.1 m**	**G.723.1 a**	**G.729**
Скорость передачи, кбит/с	64	6,3	5,3	8
Длительность кадра, мс	5	30	30	10
Задержка пакетизации, мс	1	67,5	67,5	25
Полоса пропускания для двунаправленного соединения, кГц	174,4	43,73	41,6	62,4
Задержка джиттер-буфера, мс	2-4	60	60	20
Значение R-фактора	93,2	78,2	74,2	82,2
Теоретическая максимальная оценка MOS	4,4	3,87	3,69	4,07

При этом влияние сети достаточно сильно сказывается на общей оценке и зависит от используемого кодека (рис. 16, 17), что позволяет реализовывать процедуру выбора кодека в зависимости от качества канала. Такая процедура, например, используется в решениях IP-телефонии на основе наложенных сетей (Skype, Google и пр.)

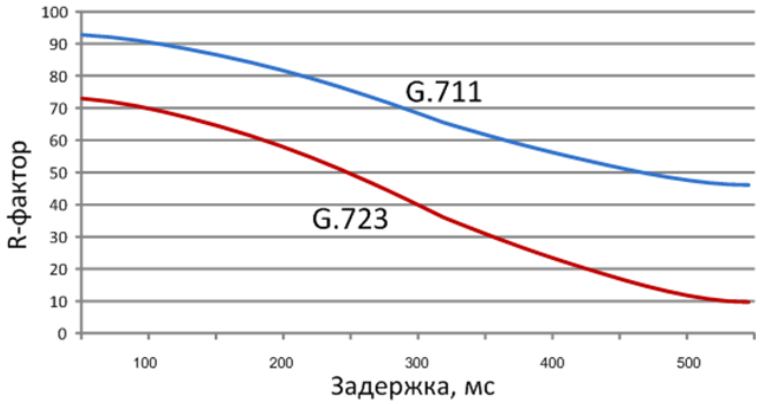


Рис. 16 – Влияние задержки на R-фактор некоторых типов кодеков

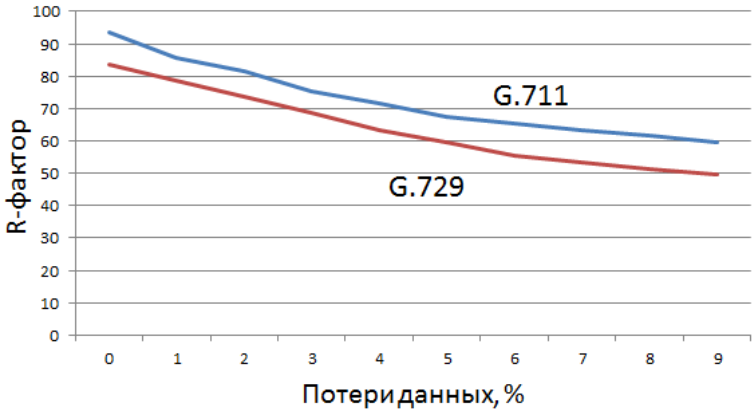
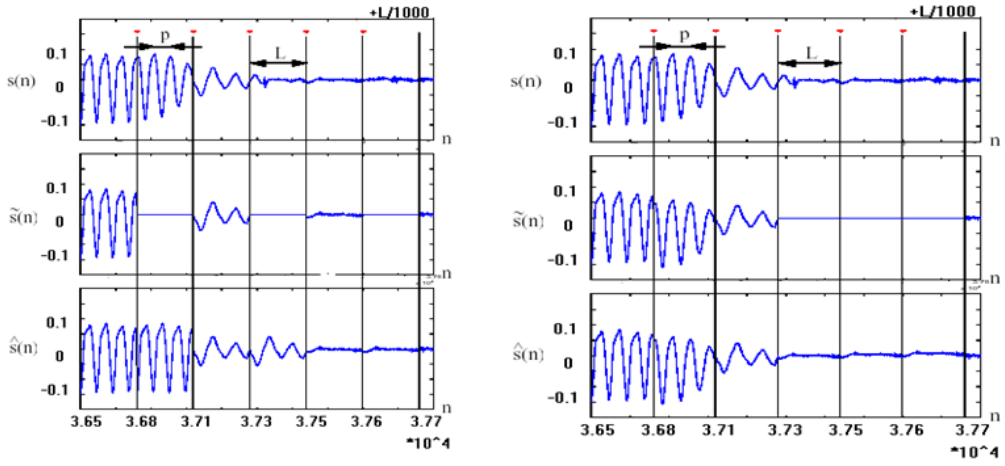


Рис. 17 – Влияние потерь на R-фактор некоторых типов кодеков

При прохождении по IP-сети за счет асинхронности появляется джиттер задержки, который может быть довольно значительным. В целях выравнивания задержек на приеме обычно реализуется джиттер-буфер, вносящий дополнительную задержку и приводящий к потерям за счет превышения максимальной задержки. Однако именно джиттер-буфер позволяет выровнять скорость потока. К тому же эффекты потерь можно компенсировать на приеме путем замены потерянных пакетов предыдущим пакетом (рис. 18, а) или комфортным шумом (рис. 18, б).



а б

Рис 18 – Методы нивелирования потерь
а) предыдущим пакетом, б) комфортным шумом

2.2 Video over IP

Сейчас происходит вытеснение эфирного телевидения Интернет-ресурсами. Такой подход привлекателен для пользователей по многим причинам: это и больший выбор, и гибкость по времени, и интерактивность. Также получили распространение Интернет-кинотеатры и каналы с возможностью прямых трансляций событий. Версии видеоконтента в IP-сетях можно разделить на следующие типы:

- VoD – Video on Demand, видео по запросу. Использует видеосервер, с которого по запросу пользователя на его терминал организуется просмотр ролика или фильма. Использует общедоступную сеть провайдера IPTV.
- Internet-TV (Streaming TV) – возможность просмотра видео через Интернет. Использует общедоступную сеть.
- IPTV – телевидение поверх IP. Один из вариантов построения сетей кабельного телевидения, в качестве протокола сетевого уровня использующего IP. Использует специализированные каналы.

Типовая схема организации IPTV представлена на рис. 19.

Основные компоненты комплекса IPTV включают в себя:

1. Головная станция (HeadEnd):
   - Антенный пост: обеспечивает прием сигналов от эфирных станций и спутников;
   - Цифровые спутниковые приемники – дескрипторы: обеспечивают раскодирование цифровых сигналов, полученных с Антенного поста и передачу материалов стримеру/мультиплексору;
   - Узел цифрового кодирования: обеспечивает MPEG-кодирование аналоговых и цифровых сигналов и передачу материалов стримеру/мультиплексору;
     - Стример/мультиплексор: обеспечивает мультиплексирование материалов и IP-вещание с присовением каждый канал уникального адреса и порта IP-вещания.

Для удобства головная станция и узел кодирования территориально находятся на одной площадке и осуществляют прием, преобразование сигналов от различных источников и формирование потоков IP-multicast/IP-unicast. Оборудование захвата контента в реальном времени получает аудиовизуальные потоки от различных источников при помощи спутниковых антенн, частотных и т.д. При необходимости шифрует, декодирует, оцифровывает и направляет их в оборудование упаковки в пакеты IP – IP-Streamer.

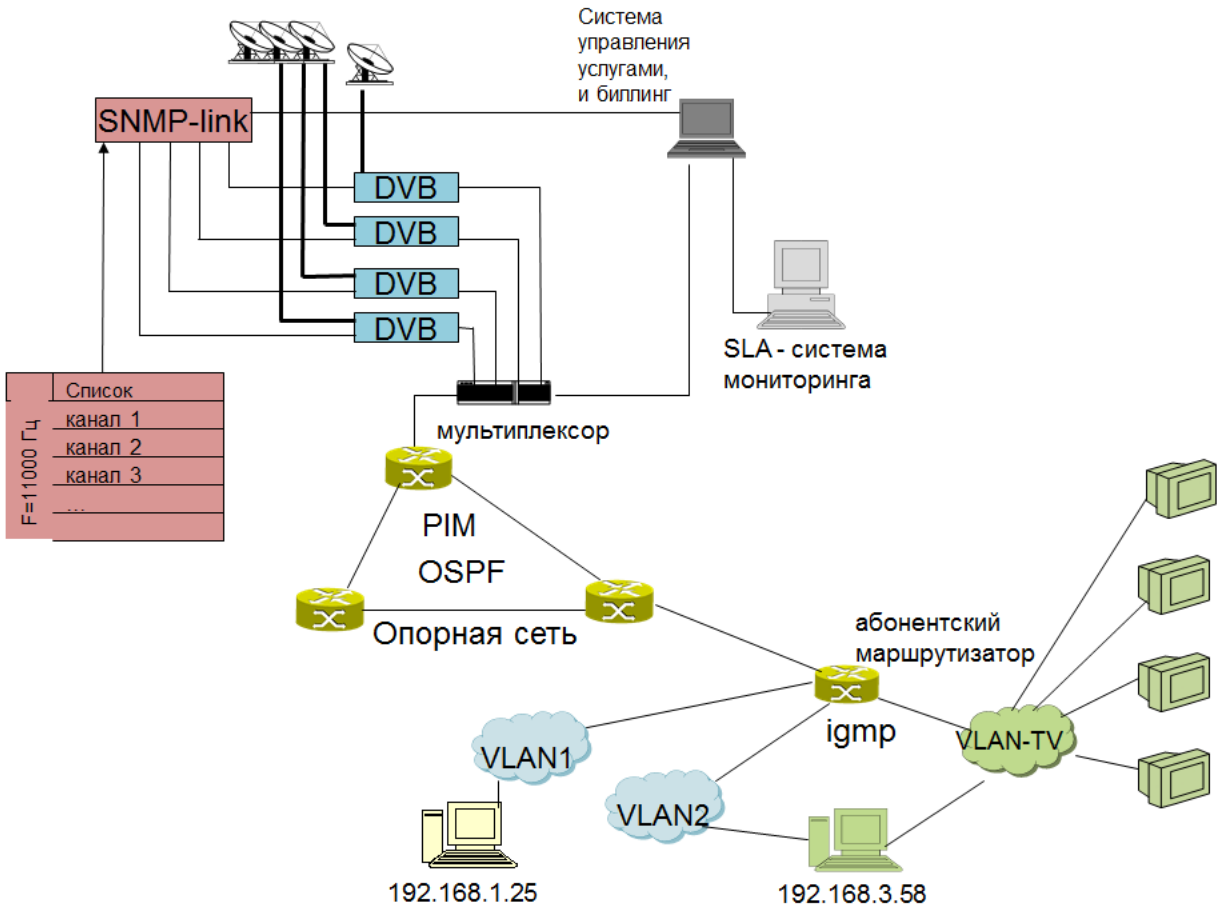


Рис. 19 – Типовая схема реализации сети IPTV

2. Подсистема видео по требованию (Video on Demand):
   - Центральный узел: аппаратно-программный комплекс, расположенный в непосредственной логической близости к системе условного доступа и головной станции.
   - Система управления: программное обеспечение, управляющее видеосерверами и распределением контента.
   - Видеосервер: аппаратно-программный комплекс, устанавливаемый для обслуживания групп пользователей на определенной территории, например, в пределах одного узла агрегации.

В задачу подсистемы VOD входит запись и проигрывание по запросу пользователя видеоматериалов.

3. Сервисная платформа (Middleware): программно-аппаратный комплекс, который обеспечивает управление всеми компонентами решения IPTV, а также служит для упрощения задач администрирования и управления предоставляемыми услугами.
4. Абонентское оборудование (STB, SetTopBox): абонентское устройство, является связующим звеном между системами формирования, доставки аудио- и видеоматериалов, и телевизором абонента, представляет собой миникомпьютер с операционной системой, web-браузером, Mpeg декодером. Может быть интегрировано в телевизионный приемник.
5. Подсистема условного доступа (Conditional Assess System): техническое средство защиты аудиовизуальных и других сообщений и материалов, распространяемых в составе ТВ-программы, позволяет разграничивать доступ пользователя к мультимедийным услугам, соблюдать авторские права, обеспечивать защиту контента от несанкционированного доступа/копирования.

Для организации вещания используется multicast - многоадресная передача, подразумевающая пересылку пакетов только тем устройствам, которые выполнили соответствующий запрос igmp, тем самым присоединившись к группе. Групповые адреса относятся к классу D, диапазон 224.0.0.0 – 239.255.255.255. Отметим, что 224.0.0.0 – 224.0.0.255 зарезервированы для протоколов маршрутизации (например, 224.0.0.5 и 224.0.0.6 для маршрутизаторов OSPF, 224.0.0.9 для маршрутизаторов RIP).

К специализированным протоколам для организации вещания поверх IP относятся:

- IGMP - Internet Group Management Protocol. Интегрируется в IP на сетевом уровне. Позволяет маршрутизатору определять принадлежность хостов к группе. Ориентирован на минимизацию служебного трафика.
- PIM - Protocol Independent Multicast MIB for IPv4. Позволяет строить покрывающее дерево в группе, причем между двумя хостами существует только один путь.
    1. PIM-DM (Dense Mode) – уплотненный режим. Используется для работы в сетях, где пользователи расположены плотно, задержки небольшие, отсутствует дефицит пропускной способности (рис. 20). PIM-DM обеспечивает гарантированную доставку, не предусматривает методов уменьшения нагрузки на сеть, использует метод широковещания и отсечения (пересылка широковещательных сообщений прекращается только после получения явного запроса на отсечение). Для маршрутизации используется любой протокол маршрутизации (чаще всего OSPF). Кратчайший путь вычисляется к каждому получателю.
    2. PIM-SM (Spase Mode) – разреженный режим. Рассчитан на работу в сетях с небольшой плотностью пользователей и ограниченными ресурсами (рис. 21). PIM-SM использует метод управления по запросу:

– Определяется точка сбора RP (Rendezvous Point), в которую отсылаются сообщения о присоединении. Маршрутизатор RP называется центральным.
– При получении IGMP запроса, локальный маршрутизатор отсылает центральному одноадресный запрос о присоединении.
– Все маршрутизаторы, находящиеся на маршруте, анализируют этот запрос о присоединении, и любой из них может ответить на запрос, если является частью дерева.

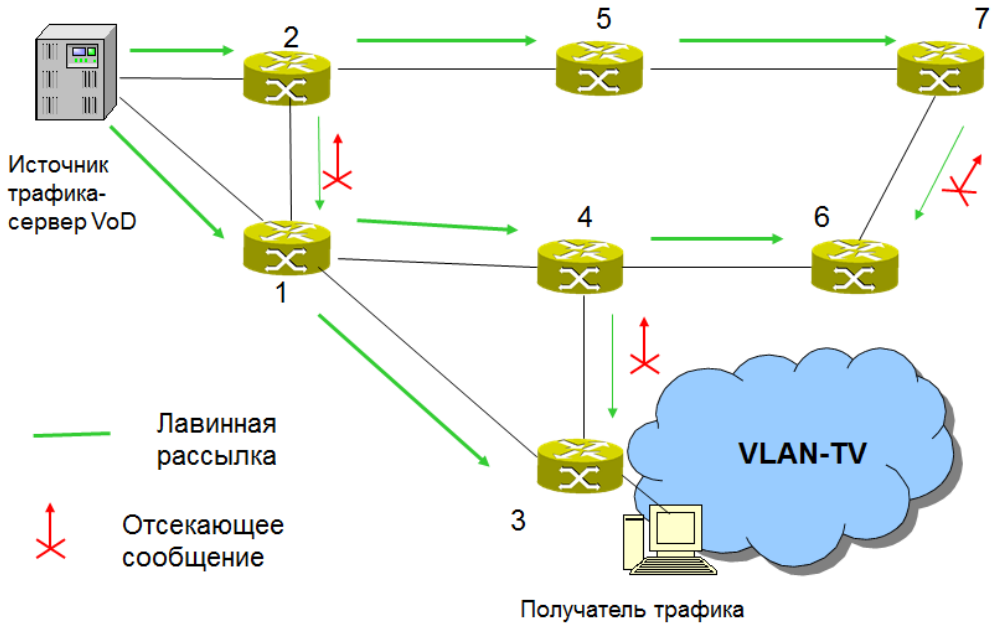


Рис. 20 – Дерево PIM-DM

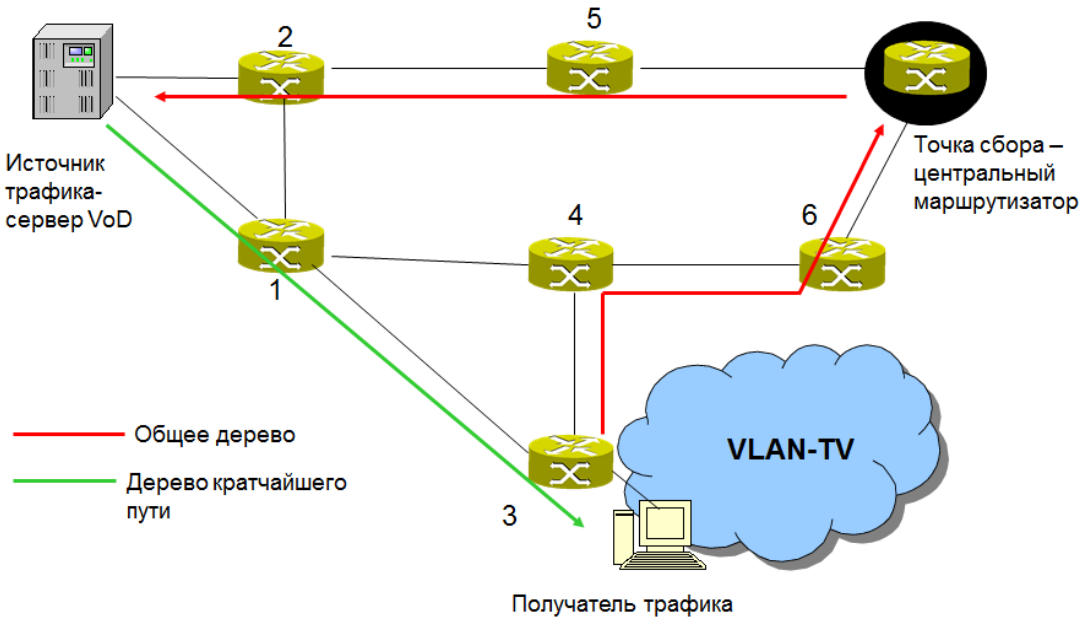


Рис. 21 – Дерево PIM-SM

PIM-SM оптимизирует возможность подключения к точке сбора с помощью процедур реконфигурации. Структура дерева может быть перестроена в случае недоступности центрального маршрутизатора или при наличии нескольких альтернативных точек сбора. Т.е. происходит переключение с общего дерева на дерево кратчайшего пути. Протокол также содержит механизм, позволяющий переключаться между общим деревом и деревом кратчайшего пути. Например, в качестве критерия переключения может служить интенсивность трафика, но тогда в силу высокой пачечности трафика требуется отдельная процедура усреднения интенсивности.

Кодеки Vedeo over IP

Базовые кодеки, используемые при предоставлении услуг:

- MPEG 1 – использует полное кодирование, требует высокоскоростных каналов.
- MPEG 2 – использует двунаправленное предсказание, кодирует полностью кадр.
- MPEG 4 – использует раздельное кодирование для разных типов видеоинформации.

MPEG 7 представляет собой не кодек, а мультимедиа-интерфейс для описания содержимого, стандартизирует элементы, ориентированные на поддержку мультимедиа. Использует понятие «уровня абстракции», виды информации о файле: форма, условия доступа, классификация, локализация, связи и т.п.

В MPEG 2 существует три типа кадров: I-кадры (Intra), P-кадры (Predicted – кадры предсказания) и B-кадры (двунаправленные). До появления кодека DivX версии 5.0 использовались только I и P кадры. I-кадры используют информацию только из самого кодируемого кадра, основаны на сжатии одиночных кадров в формат JPEG. P-кадры (Predicted – кадры предсказания) предсказывают следующие кадры и могут также ссылаться на I- или P-кадры, т.е кодируются с использованием информации из предыдущих кадров. В любой видеопоследовательности всегда найдется группа кадров, которые будут содержать одно и то же изображение. Таким образом, вместо независимого JPEG-кодирования каждого кадра, можно использовать избыточность предыдущих кадров – передается только разность между соседними кадрами.

B-кадры строятся на анализе предсказания кодеком не только будующих кадров, но и кадров предсказанных ранее, такими как I или P-кадры. Использование сокращает объем данных, требуемых для кодирования кадра. Кроме того, использование B-кадров улучшает качество фильма, особенно в областях кадра, где движущиеся объекты открывают скрытые области. С B-кадрами связано наиболее глубокое сжатие видеоданных. Поскольку высокая степень сжатия снижает точность восстановления исходного изображения, B-кадры не используются в качестве опорных.

Точность кодирования должна быть максимальной для I кадров, ниже для P кадров и минимальной для B кадров.

Обработка видеоданных в P-кадре выполняется по макроблокам. Это квадратные матрицы 16 x 16 (отсчетов × строк). Такой макроблок обрабатывается с использованием алгоритмов компенсации движения и предсказания вперед пока

в блоке не появится новый объект. С этого момента процесс кодирования переключается на алгоритмы, используемые в I-кадрах. P-кадры являются опорными для последующих P- или B-кадров. Ошибки опорного кадра распределяются по всем кадрам, связанным с опорным. Ошибки при их декодировании не распределяются по другим кадрам.

Алгоритмы кодирования B-кадров зависят от характера ТВ-изображения. Предусмотрено четыре способа кодирования.

1. Применяется компенсация движения и предсказание вперед по ближайшим предшествующим опорным I- или P-кадрам.

2. Компенсация движения и обратное предсказание по ближайшим последующим I- или P-кадрам. Обратное предсказание используется в тех случаях, когда в кодируемом B-кадре появляются новые объекты изображения.

3. Компенсация движения и двунаправленное предсказание, при котором опорными являются предшествующий или последующий I- или P-кадры.

4. Внутрикадровое предсказанием без компенсации движения. Такое кодирование нужно при резкой смене передаваемых сюжетов, а также при больших скоростях перемещения объектов ТВ-изображения. Ошибки при их декодировании не распределяются по другим кадрам.

Точность кодирования должна быть максимальной для I кадров, ниже для P кадров и минимальной для B кадров.

Для уменьшения задержки на кодеке используется чередование кадров (рис. 22). Типичным является порядок кодирования I, P, B кадров: группы, состоящие из 12 чередующихся кадров: I0, B1, B2, P3, B4, B5, P6, B7, B8, P9, B10, B11, I12, B13, B14, P15 и т. д., в которых I кадры следуют с интервалом: (1/25 Гц)×12= 0,48 с. При передаче по каналу связи порядок следования I, P и B кадров меняется. В декодер в начале поступают опорные I и P кадры, без которых нельзя начать декодирование. Типичным является следующий порядок передачи I, P, B кадров: I0, P3, B1, B2, P6, B4, B5, P9, B7, B8, I12, B10, B11 - P15, B13 и т. д.

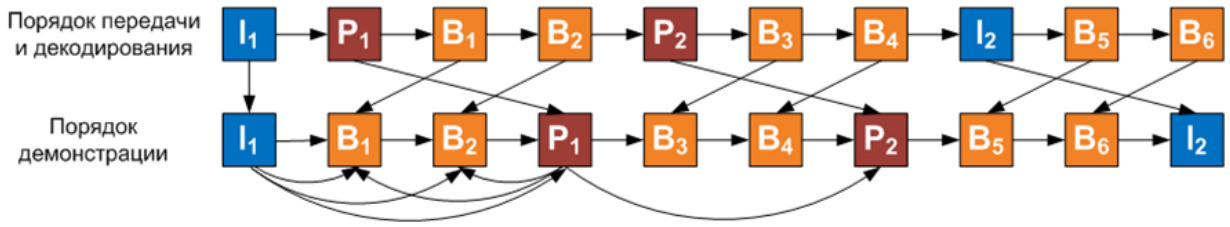


Рис. 22 – Чередование кадров

Формирование видеопотока (на рис. 23) подразумевает создание специальных заголовков на каждом уровне для синхронизации видео.

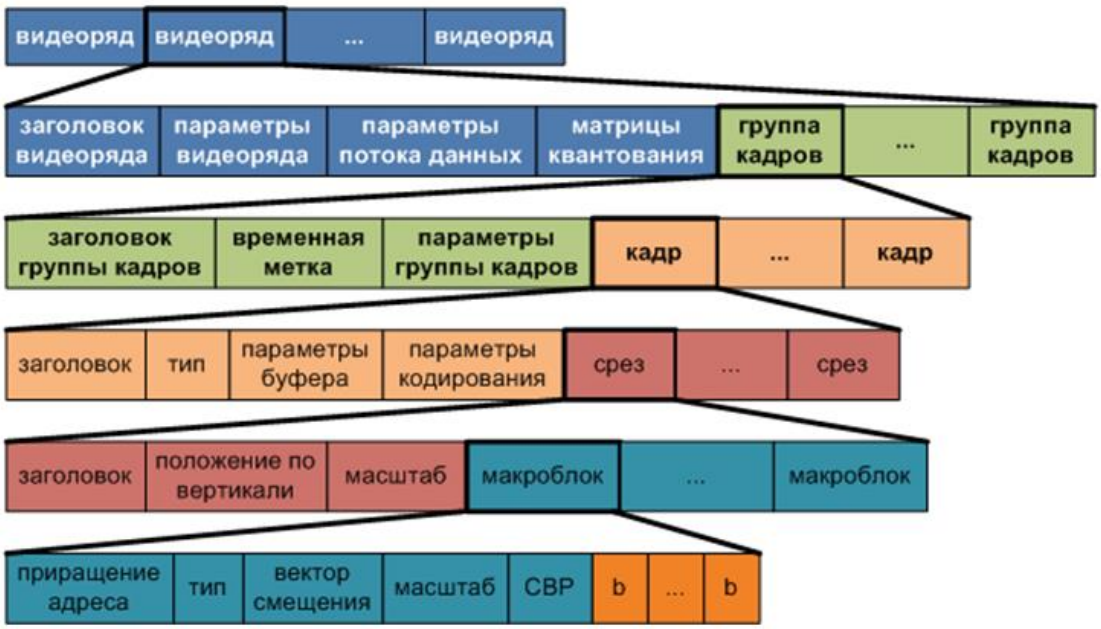


Рис. 23 - Структура видеопотока

После формирования видеопотока происходит формирование транспортного потока для передачи по сети (рис. 24).

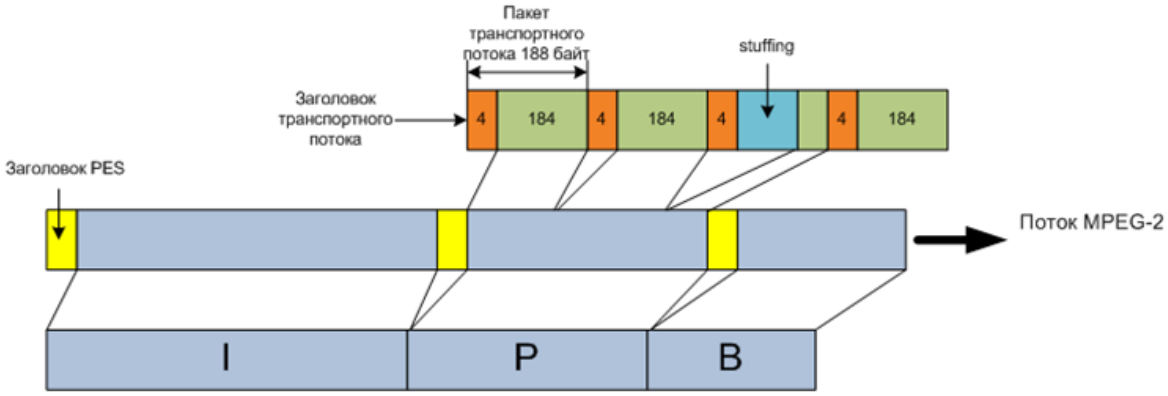


Рис. 24 – Формирование транспортного потока

После формирования транспортного потока начинается формирование IP-пакетов (рис. 25).

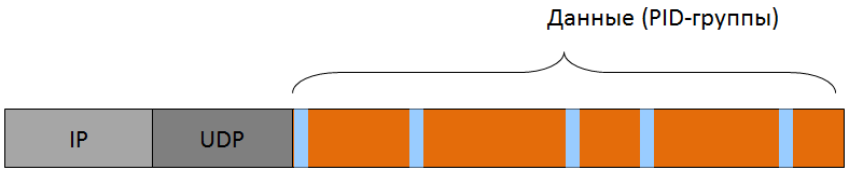


Рис. 25 – Формирование IP-пакета с видеопотоком

Приняты следующие обозначения:
PID – идентификатор потока;
PES – элементарный поток пакетов: пакеты видео и аудио данных неопределенной длины;
PCR – блок синхронизации, передается каждые 0,1 с., частота 27 МГц.

Можно отметить следующие особенности кадров видео на сетевом уровне:

- Заголовок протокола прикладного уровня отсутствует.
- В одном кадре передается информация о звуке, видео, тексте.
- Размер кадра соответствует MTU=1500 байт.
- Используемые типы видео кодеков: MPEG2 и MPEG4.
- Стандарт MPEG4 разработан с опорой на MPEG2, поэтому, несмотря на разницу формирования кадров в MPEG2 и MPEG4, алгоритм формирования потока трафика сохраняется.

Таким образом, для передачи по пакетной асинхронной сети синхронного потока видео используется механизм формирования заголовков с параметрами синхронизации на всех уровнях, от кодека до транспортного потока, до сетевого уровня. Такой подход позволяет разгрузить сеть от необходимости проверять синхронизацию и порядок следования кадров. Как недостаток можно отметить большое количество служебной информации.

Оценка качества видеоуслуг

Для оценки видео используется Quality of Experience, QoE (G.1011, G.1080), которое отражает степень удовлетворённости пользователей

Методология измерения субъективного восприятия качества телевизионного изображения изложена в Рекомендации ITU-R BT.500-11. Можно выделить следующие этапы тестирования:

- Выбор или сочетание способов демонстрации видеопоследовательностей.
- Определение методики сбора мнений экспертов.
- Выбор методики обработки результатов.

Субъективные методы оценки качества видео:

1. Mean Opinion Score, MOS. Количественная оценка качества видеоуслуги производится по следующим параметрам экспертом:
   - блочность (распад изображения на квадраты);
   - смазанность;
   - кадровое дрожание (Jerkiness);
   - случаи остановки кадров;
   - случаи потери кадров;
   - кратковременные сбои (Temporal complexity);
   - пространственные сбои (Spatial complexity).
2. Difference Mean Opinion Score, DMOS - разница между оценками MOS тестируемого и эталонного видео.
3. Picture Quality Rating, PQR (BT.500-11) - определяет способность пользователя отличать эталонное видео и тестируемого фрагмента.

$$видео_{тест.} = видео_{эт.} + k \times (видео_{ухуд.} - видео_{эт.}),$$

где $0 < k < 1$ – коэффициент улучшения, определяется в процессе оценки видео экспертами.

4. Subjective Assessment Method for Video Quality evaluation, SAMVIQ – субъективная оценка качества после компрессии кодеком, для калибровки шкалы используется тестовый видеоряд с выбранным экспериментатором кодеком.

Объективные методы оценки качества видео:

1. Media Delivery Index (MDI, RFC 4445, 1996): оценивает влияние задержки, джиттера задержки, потерь без учета особенностей видео, может использоваться на любом участке сети.
2. Moving Pictures Quality Metric (MPQM, V-фактор, 1996): анализирует контрастность, размытость, рассыпание изображения, замирание, нарушение цветности, артефакты по 6-бальной шкале по 34 параметрам.
3. Video Quality Metrics (VQM, BT.1683, 1999): анализирует смазанность (размытие), дрожание, блочность, шум, искажение цвета, приводя к единой метрике.
4. Noise Quality Measure (NQM, 2000) оценивает влияние аддитивного шума на исходный сигнал, анализируя контрастность и яркость. Оценивает влияние потерь.
5. Peak Signal-to-Noise Ratio (PSNR) – популярная метрика, пиковое отношение сигнал/шум. Оценка может использоваться как приблизительная, т.к. не дает гарантию, что зрителю понравится восстановленный образ, хотя имеет связь с MOS (табл. 6). Оценивает только влияние потерь.

Таблица 6 - Связь между PSNR и MOS

MOS	1 Очень плохое	2 Плохое	3 Среднее	4 Хорошее	5 Очень хорошее
PSNR	<20	20-25	25-31	31-37	>37

Отметим, что некоторые из методов хорошо отслеживают потери, некоторые – задержки, но ни один из методов не может оценить влияние джиттера задержки и сбоя синхронизации на верхних уровнях (табл. 7).

Рассмотрим основные проблемы реализации видеоуслуг на базе существующих сетей:

- Слабая поддержка абонентской сетью: пропускные способности на абонентском участке должны иметь возможность поддерживать по крайней мере два видеопотока одновременно для возможности переключения каналов.
- Недостаточный ресурс сегмента транспортной сети: при большом количестве пользователей возрастает количество каналов, просматриваемых в одно время, а также другие типы трафика, в том числе услуг OTT.
- Не реализовано обеспечение гарантированного качества обслуживания.

– Управление сервисами (высокие значения задержек при подключении сервисов, переключении каналов, изменении параметров сервисов и пр.).
– Необходимость разработки видеоконтента с учетом запросов пользователей.

Таблица 7 – Зависимость искажений видео от сетевых показателей [7]

Искажение видео	**Сетевые показатели**	**Метод измерения**
Блочность видео (пикселизация)	Потери пакетов	VQM, PSNR, MDI, MPQM
Случайные полосы в изображении	Джиттер	MDI
Искаженное видео, рассыпание изображения	Прибытие пакетов в неправильном порядке	
Несинхронный звук по отношению изображению	Ошибки в настройке буфера на приеме, джиттер	
Смазанность	Уровень потерь 5%	VQM, PSNR, MDI, MPQM
Мерцание	Переполнение буфера абонентского устройства и потери пакетов	
Застывшие видеокадры	Разный уровень потерь	VQM, PSNR, MDI, MPQM
Провалы в видеоизображении		
Отсутствие видео		
Потеря звука в видео	Задержки, джиттер	MDI

Сформулируем основные требования к сети IPTV:
– Пропускная способность на абонентском участке минимум 24 Мбит/с (с учетом взрывного характера трафика).
– Пропускная способность VLAN от 100 Мбит/с.
– Поддержка мультикастового (multicast) трафика при организации IPTV.
– Поддержка одновременной передачи трафика различных типов (эластичного, потокового и реального времени) транспортной сетью.
– Поддержка стандартов кодирования MPEG2 и MPEG4.
– Поддержка QoS согласно заключенному соглашению об уровне обслуживания.

3 ФОРМИРОВАНИЕ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ОБСЛУЖИВАНИЯ В IP-СЕТЯХ

Как уже отмечалось, показатели QoS согласно Y.1540 можно разделить на два типа. К показателям, характеризующим проблемы передачи трафика относятся:
– Задержка в сквозном соединении и джиттер* задержки (мс).
– Величина потерь (%).

К показателям, характеризующим возможности сети относятся:

– Производительность сети (бит/с).
– Надёжность сетевых элементов (Кг).
– Устойчивость (живучесть) функционирования сети - возможность сохранения функционала при выходе из строя отдельных элементов (резервирование оборудования по схеме n+m)

Рассмотрим поведение трафика и формирование задержек, потерь и джиттера задержек в асинхронной IP-сети.

3.1 Классификация трафика мультисервисной сети

Трафик мультисервисной сети можно представить потоками трех основных типов (табл.8). Первый поток – это так называемый *эластичный трафик*, т.е. такой, что изменение пропускной способности участка сети почти не сказывается на качестве обслуживания. Эластичный трафик чувствителен к потерям и не чувствителен к задержкам. Второй – *потоковый трафик*, допускающий достаточно большие задержки (до 700мс). И, наконец, третий – трафик, чувствительный к задержкам и относительно малочувствительный к потерям, т.е. трафик IP-телефонии и видеоконференцсвязи, так называемый *трафик реального времени*.

Таблица 8 - Концепция Triple Play [7]

Трафик	Приложения	Чувствительность к показателям качества
Реального времени	IP-телефония, видеоконференцсвязь	К задержке К джиттеру задержки Малая к потерям
	управление, транзакции, игры on-line	К задержке К джиттеру задержки К потерям
Потоковый	Аудио по требованию, видео по требованию, Интернет-вещание	Малая к задержке К джиттеру задержки К потерям
Эластичный (трафик передачи данных)	Документооборот, управление БД	Малая к задержке Малая к джиттеру задержки Высокая к потерям
	Анимация, передача данных датчиков, передача файлов, E-mail, веб	Очень малая к задержке Малая к джиттеру задержки Высокая к потерям

Эластичный трафик – трафик передачи данных, в качестве транспортного протокола использующий TCP. Этот трафик отличается низкой чувствительностью к задержкам (до нескольких минут в зависимости от приложения) и не допускает потерь в сквозном соединении. Примерами это вида трафика является трафик, генерируемый такими приложениями, как e-mail, пересылка файлов, web-приложения и т.п.

Потоковый трафик порождается такими приложениями, как Интернет-вещание, аудио и видео по требованию. Этот тип трафика малочувствителен к

потерям, малочувствителен к задержкам и джиттеру задержки. На приеме обычно используется джиттер-буфер, позволяющий сглаживать неравномерность задержки путем внесения дополнительной задержки буфера. Таким образом, для передачи этого типа трафика вполне возможно использование в качестве транспортных протоколов как UDP, так и TCP.

Приложения реального времени можно разделить на два типа: трафик транзакций и трафик реального времени, соответствующий мультимедийным приложениям. Оба типа трафика реального времени характеризуются высокой чувствительностью к задержкам и джиттеру задержки. В зависимости от класса обслуживания оговариваются их конкретные значения. Трафик транзакций представляет собой сигналы управления различными объектами и процессами, в том числе игры on-line. Такой тип трафика предъявляет высокие требования к задержке, т.е. относится к сверхчувствительному к задержкам типу, характеризуется высокой чувствительностью к потерям и переменной битовой скоростью (т.е. отличается высокой степенью непредсказуемости).

Трафик реального времени, порожденный такими процессами как речь или видео, отличается большей устойчивостью к потерям (т.е. относится к малочувствительным к задержкам типам приложений), является изохронным (т.е. имеет порог чувствительности к задержкам, при превышении которого функциональность приложения резко снижается), характеризуется высокой степенью предсказуемости порождаемого трафика.

Таким образом, в мультисервисной сети можем наблюдать различные комбинации этих трех видов трафика. В силу различия приложений и, следовательно, требований к качеству обслуживания, каждый тип трафика и его система обслуживания характеризуется различными законами распределения. Трафик мультисервисной сети на сетевом уровне можно представить как их совокупность.

3.2 Модель формирования показателей качества в сквозном соединении

Сквозная модель из конца в конец (end-to-end или e2e) была предложена ITU-T. Эта модель ориентирована на определение QoS в наиболее важной с точки зрения конечного пользователя степени. Эталонная модель сквозного QoS обычно содержит одну или более взаимодействующих сетей, каждая из которых потенциально имеет ряд узлов (рис.26).

Каждая из этих взаимодействующих сетей может вносить свои задержку, потери или ошибки вследствие процедур мультиплексирования, коммутации или передачи, что отрицательно влияет на результирующее значение показателей QoS. Кроме того, статистические колебания поступающего трафика могут приводить к потерям вследствие переполнения буферов на каналах, соединяющих сетевые узлы, испытывающие перегрузку.

Рекомендация I.356 специфицирует QoS по наихудшему варианту соединения сетей и устройств. Таким образом, до тех пор, пока характеристики соединений между сетями и устройствами остаются в рамках этих границ, пользователи получают надлежащий согласованный уровень QoS.

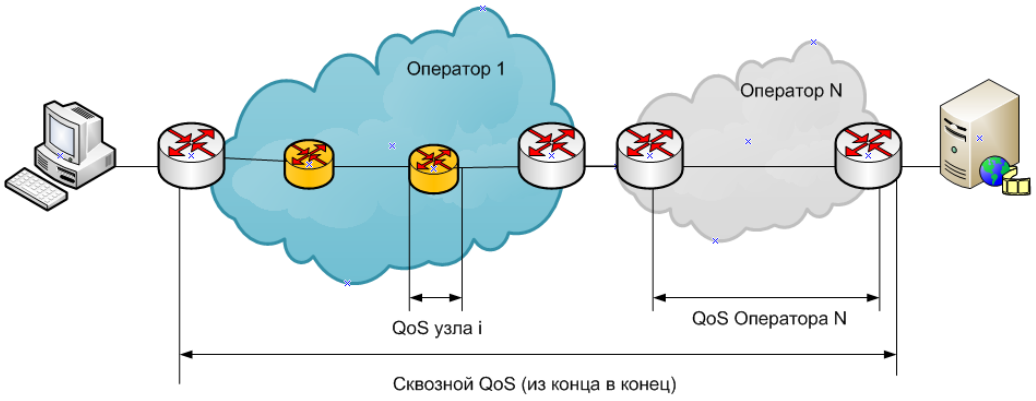


Рис.26 - Эталонная модель сквозного QoS

На представленной на рисунке 26 эталонной модели сквозного качества обслуживания, влияние на качество обслуживания речевого трафика оказывают:

– оконечные устройства на обоих концах могут влиять на четкость передачи речи и видео из-за качества кодека, камеры, микрофона, а также возможности некомпенсированного акустического эха;
– элементы сети (узлы, каналы) влияют на качество передачи сигнала из-за задержек, джиттера и потерь пакетов.

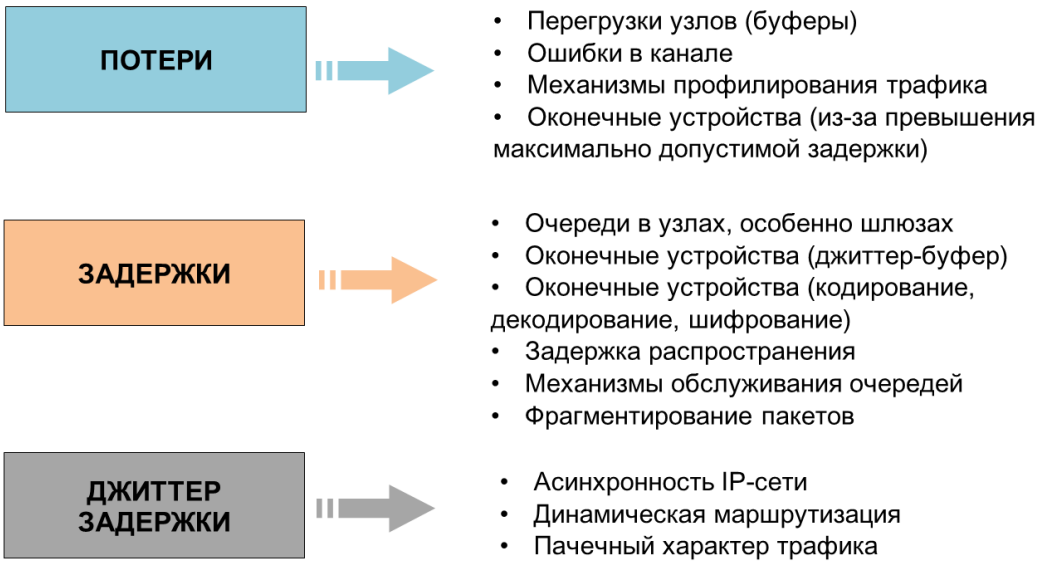


Рис. 27 - Показатели QoS, учитываемые при передаче мультимедийного трафика и механизмы их формирования

В качестве примера влияния устройств сети на межконцевую задержку рассмотрим трассировку к удаленному узлу, находящемуся в другом регионе. На основании трассировки построим график зависимости времени от количества пройденных узлов (рис. 28). Красными стрелками помечены узлы, являющиеся шлюзами и вносящие постоянную задержку.

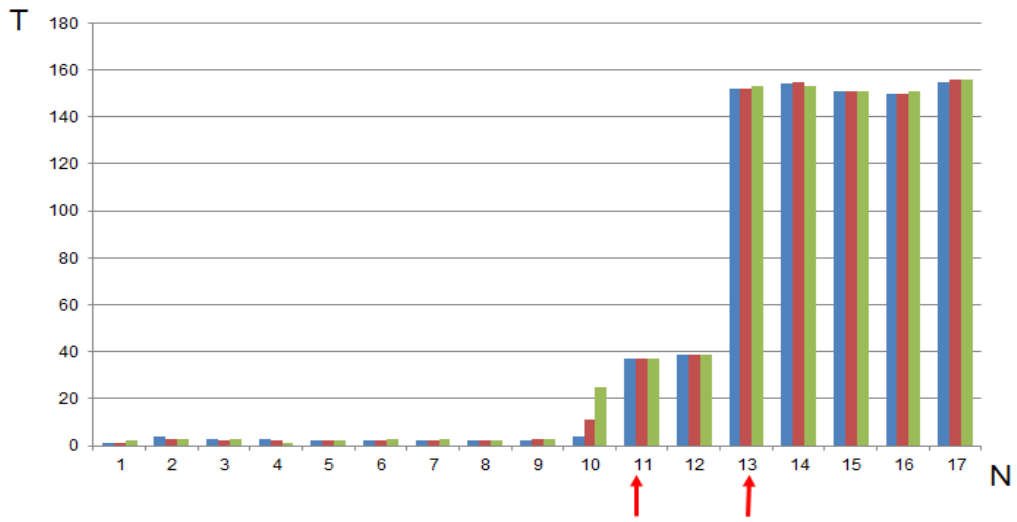


Рис. 27 – Диаграмма времени трассировки для узла 141.8.225.72 из Санкт-Петербурга

Суммарная (межконцевая или end-to-end) задержка определяется как сумма задержки оконечного устройства и сетевой задержки.

Задержки оконечного устройства включают:

– задержку на обработку фрагмента речевого сигнала на передающей и принимающей сторонах, т.е. время, требуемое для кодирования и декодирования сигнала, и в основном зависят от алгоритма работы кодека;
– задержку джиттер-буфера - задержки на принимающей стороне для компенсации разброса во временах прибытия речевые пакетов (джиттера);
– задержку формирования пакетов, зависящую от выбранной схемы формирования и определяется алгоритмической задержкой кодека, периодом формирования и количеством речевых кадров в IP-пакете.

Сетевые задержки определяются ожиданием в очередях и задержками маршрутизации в маршрутизаторах.

Формирование задержки и джиттера задержки

Рассмотрим подробнее формирование задержек в узле.

Для начала обратим внимание, что трафик современных мультисервисных сетей описывается тяжелохвостыми распределениями. Обратимся к основам теории телетрафика. Принята следующая терминология:

- входящий поток вызовов (требований на обслуживание) – стохастический процесс, поток однородных событий.
- система распределения и обработки информации – сеть связи вообще или какой-то из ее элементов, например, узел (рис. 29)
- дисциплина обслуживания потока вызовов - описывает взаимодействие потока вызовов с системой распределения информации.

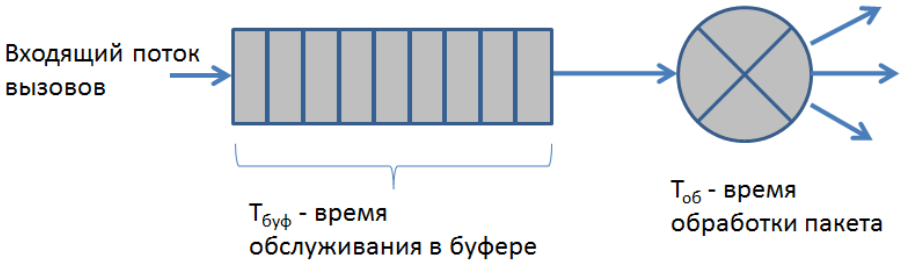


Рис. 29 – Узел с точки зрения теории телетрафика

Для описания процессов в узле используют модель Кендалла A/B/m/k/M, где:
A - распределение промежутков времени между последовательными требованиями
B - распределение времени обслуживания требований
m – число обслуживающих приборов
k - длина очереди
M – число источников нагрузки.

Законы, описывающие распределение согласно A и B, могут быть различными. В общем виде принято обозначение G. Для описания процессов в телефонной сети общего пользования традиционно пользовались законом Пуассона (обозначается как M). Однако при переходе к пакетным сетям появляется корреляционная зависимость, связанная с образованием пачечности, и более адекватное описание дают так называемые тяжелохвостые зависимости:
W – Вейбулла-Гнеденко
LN – логнормальное
P – Парето

Такие процессы получили название самоподобных. Причины возникновения самоподобия в мультисервисных сетях:

- Поведение пользователя: довольно часто пользователи почти синхронно используют ресурсоемкие приложения или услуги OTT (Over The Top – видео, передаваемое по наложенной сети). Это приводит к образованию трафика с высокой пачечностью, что, соотвественно, порождает кратковременные перегрузки на узлах.
- Способ генерации трафика: структура видеопотока, компрессия речи, закачка файлов – все эти факторы также приводят к выраженной пачечности.
- Агрегирование потоков: на устройствах агрегации объединение нескольких потоков с высокой пачечностью усиливает эффект и приводит к образованию еще более тяжелохвостых зависимостей. Наоборот, методы профилирования трафика позволяют сглаживать неравномерность агрегированного потока.
- Использование механизмов управления и образование обратной связи: перезапросы, медленный старт и механизмы обслуживания очередей могут послужить причиной возникновения очередей в буферах и, как следствие, неравномерности увеличению джиттера задержки в потоке денных.

Конечно, такие изменения в структуре трафика повлекли за собой изменения в оценках основных показателей качества обслуживания.

Рассмотрим задержки и их формирование в сквозном соединении.

Для пакетного трафика можно рассматривать общую задержку t (или время доставки пакета) как сумму транспортной задержки t_{tr}, задержки распространения t_p, задержки коммутации t_s и задержки при организации очередей в маршрутизаторах (времени задержки в узле).

$$t = t_{tr} + t_p + t_s + t_{□}.$$

Под *транспортной задержкой* (serialization delay) подразумевается время, требуемое для передачи пакета при заданной полосе пропускания, и зависит от размера пакета и ширины полосы пропускания канала. Транспортную задержку можно определить как функцию от ширины полосы пропускания и длины пакета, т. е.

$$t_{tr}=L/B,$$

где L – размер пакета, бит, B – ширина полосы пропускания, кбит/с.

Задержка распространения (propagation delay) зависит от используемой среды передачи и расстояния и может составлять десятки миллисекунд. Внедрение технологии DWDM позволяет принимать задержку распространения менее 1 мс.

Задержка коммутации (switching delay) вносится устройствами коммутации и, как правило, составляет менее 10 нс.

В случае, если сеть не испытывает перегрузки, задержка при организации очередей в маршрутизаторах t_ω отсутствует. В этом случае можно говорить о минимально возможной задержке при передаче пакетов через заданную сеть. В случае перегрузки сети t_ω не только может составить значительную величину, но и приводит к джиттеру задержки. Джиттер задержки и определяет максимальную задержку на приеме. В зависимости от типа приложения, принимающая сторона может попытаться компенсировать джиттер задержки за счет организации приемного буфера для хранения принятых пакетов на время t_j, меньше и равное верхней границе дрожания.

Для трафика реального времени джиттер задержки может привести к потере пакетов, т.к. при превышении порогового значения задержки пакеты будут отброшены как не удовлетворяющие требованиям, предъявляемым к режиму реального времени. Для потокового трафика внесение дополнительной задержки не оказывается критичным и не приводит к потерям.

Значение задержки t_{delay} в узле могут быть определены как:

$$t_{delay} = \bar{t}_\omega + \bar{t}_s ,$$

где

$\bar{t}_\omega = P(\rho, m)\frac{\bar{t}_s}{m(1-\rho)} \cdot \frac{C_a^2 + C_s^2}{2}$ - среднее время пребывания пакета в буфере,

$\bar{t}_s$ - среднее время обработки пакета прибором обслуживания

$C^2 = \left(\frac{\sigma}{M[x]} \right)^2$ - квадратичный коэффициент вариации.

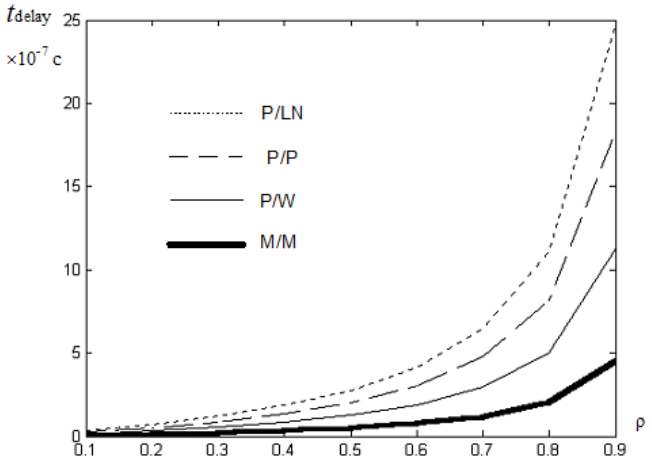


Рис. 30 - Зависимость величины задержки от ρ
для различных законов распределения входного потока

Таким образом, на задержку трафика влияют закон распределения трафика, максимальный размер пачек пакетов и минимальный интервал между приходами пачек.

Формирование потерь

Потери сети P_{net} обусловлены ошибками в канале (с внедрением ВОЛС потери в канале порядка 10^{-9} и ими можно пренебречь) и потерями в узлах сети P_{loss}. Потери в узлах сети P_{loss} определяются интенсивностью трафика, размером буфера, применяемой политикой обслуживания очередей и используемыми методами предотвращения перегрузки:

$$P_{loss} = \frac{1-\rho}{1-\rho^{\frac{2}{C_a^2+C_s^2}nb+1}} \rho^{\frac{2}{C_a^2+C_s^2}nb},$$

где C_a^2 и C_s^2 – квадратичные коэффициенты вариации соответственно распределений входящего потока и времени обслуживания, nb – размер буфера, ρ - загрузка системы, P_{loss} – потери на одном узле.

Формирование потерь в сквозном соединении может быть оценено как

$$P_{e2e} = 1 - (1 - P_{net})(1 - P_{ter}),$$

где P_{net} – потери сети, P_{ter} – потери на оконечном устройстве из-за превышения допустимой задержки (влияние джиттер-буфера) или повреждения пакетов (проверка контрольной суммы на транспортном уровне).

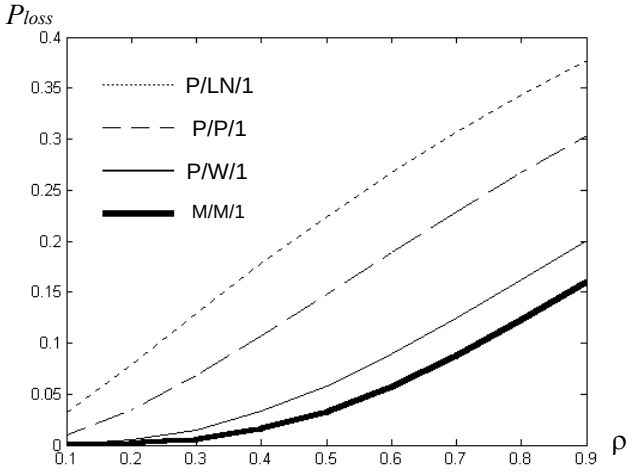


Рис.31 - Зависимость вероятности потерь от ρ
для различных законов распределения входного потока

Важно: потери сети P_{net} включают в себя потери на узлах и потери в канале и могут быть рассчитаны по аналогичной формуле:

$$P_{net} = 1 - (1 - P_{chanel})(1 - P_{hosts}),$$

При этом потери в канале ВОЛС $\approx 10^{-10}$ обычно игнорируются, потери в радиоканале могут быть существенными.

Устойчивость функционирования сети

Рассмотрим причины отказов в IP-сетях (рис. 32). Легко видеть, что основное количество отказов – более 50% - связано с маршрутизаторами: их неправильной конфигурацией, взаимодействием с другими узлами, низким коэффициентом готовности и наработки на отказ. Следующими по распространенности идут аварии на физических линиях, связанные с обрывом кабеля, его намоканием и пр. Остальные причины вносят относительно небольшой вклад в статистику отказов.

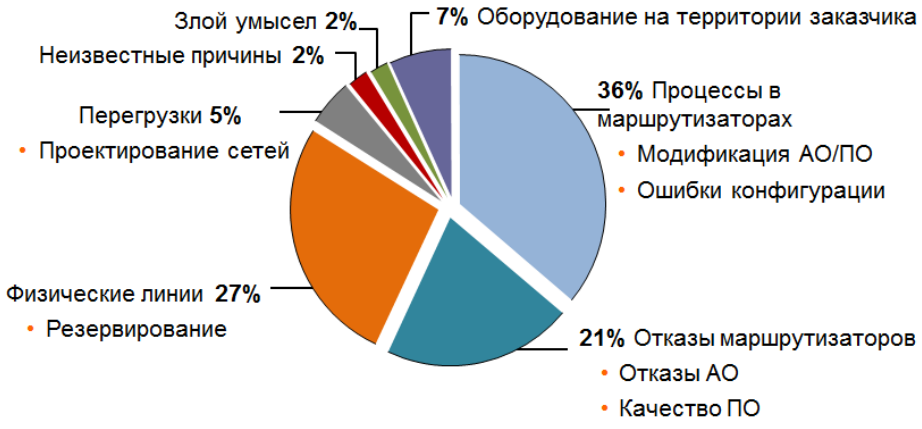


Рис. 32 - Причины отказов в IP-сетях

Важным показателем качества обслуживания является надежность функционирования сети. Оценка надежности сети производится для полученного на предыдущем этапе множества минимальных путей с помощью одного из алгоритмов и имитационного моделирования, предварительно представив сеть в виде графа. Надежность сети оценивается в заданном направлении между двумя вершинами графа (рис. 33). Т.е. надежность понимается как вероятность обеспечения связи в заданном направлении, определяемая структурой сети, правилами маршрутизации и надёжностью основных элементов**.** Показатели надежности и живучести носят вероятностный характер, так как влияние внутренних и внешних дестабилизирующих факторов показателей надежности и живучести систем являются прогнозированными.

Рис. 33 - Алгоритм вычисления оценок надежности сети

Рассмотрим причины понижения надёжности сети (рис. 34). Приблизительно 80% проблем обусловлены человеческим фактором, как прямым воздействием (ошибки эксплуатации), так и косвенным (разработка программного обеспечения).

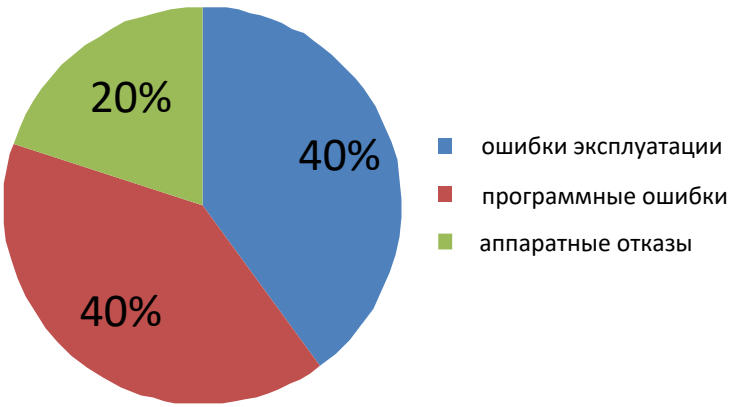


Рис. 34 - Причины понижения надежности сети

Для расчета надежности сети существует довольно большое количество методов:

- перебора простых цепей (МППЦ);
- декомпозиции;
- замещения;
- Изори-Проскана;
- включения-исключения;
- преобразования «дельта-звезда».

В основе каждого из методов лежит понятие коэффициента готовности и оперативной готовности оборудования и времени наработки на отказ.

Коэффициент надежности Кг рассчитывается по формуле [9]:

$$\text{Кг} = \frac{\text{T}_0^{(1)}}{\text{T}_0^{(1)} + \text{T}_B^{(1)}}$$

$\text{T}_0^{(1)}$ - средние значения промежутков времени между отказами.

$\text{T}_B^{(1)}$- средние значения промежутков времени между восстановлениями.

Коэффициент оперативной готовности канала связи рассчитывается как:

$$K_{\text{о.г.}} = P(t) \cdot K_{\text{г}} ,$$

где $K_г$ - коэффициент готовности; P(t) - вероятность сохранения работоспособности канала электросвязи при воздействии внешних дестабилизирующих факторов. Существует зависимость между Кг и временем простоя (табл. 9).

Таблица 9 - Коэффициенты готовности и соответствующие им значения времени простоя оборудования

Коэффициент готовности		**Время простоя**
0,99	“две девятки”	3,7 дней в год
0,999	“три девятки”	9 часов в год
0,9999	“четыре девятки”	53 минуты в год
0,99999	“пять девяток”	5,5 минут в год
0,999999	“шесть девяток”	30 секунд в год

Производители оборудования чаще всего используют понятие «время наработки на отказ» - Mean time between failures (MTBF) – задается в часах и публикуется в разделе «Характеристики оборудования». Это значение показывает интенсивность случайных отказов, но при этом исключает систематические неисправности, которые могут быть вызваны ошибками при проектировании, сбоями программного обеспечения или дефектами изготовления изделия в начале срока службы, исключает износ в ходе эксплуатации изделия ближе к концу срока службы.

Для повышения устойчивости функционирования сети используется несколько подходов:

1. *Резервирование* — метод повышения характеристик надёжности технических устройств или поддержания их на требуемом уровне посредством введения аппаратной избыточности за счет включения запасных элементов и связей, дополнительных по сравнению с минимально необходимым для выполнения заданных функций в данных условиях работы.

Кратность резервирования — отношение числа резервных элементов к числу основных элементов устройства.

По состоянию резервных элементов:

- резерв нагруженный — резервные элементы нагружены так же, как и основные;
- резерв облегчённый — резервные элементы нагружены меньше, чем основные;
- резерв ненагруженный — резервные элементы практически не несут нагрузки.

2. *Зеркалирование* ресурсов или DNS-балансировка - заключается в присваивании хосту с одним именем нескольких альтернативных *ip*-адресов, что позволяет распределять трафик посредством традиционной маршрутизации.
3. *Балансировка трафика* – набор методов, позволяющих распределять трафик внутри сегмента сети для предотвращения перегрузок.
4. *Виртуализация оборудования* (VRRP и др. протоколы, SDN)
5. Меры по повышению коэффициента готовности.
6. Меры по обеспечению сетевой безопасности и информационной безопасности – обеспечение защищенного физического доступа к оборудованию, защищенного удаленного администрирования, борьба с паразитным трафиком, в том числе DDOS-атаками.

4 TRAFFIC ENGINEERING

Traffic Engineering (TE) – это группа методов и механизмов, позволяющих осуществлять управление трафиком и влиять на его характеристики с целью обеспечения заданного качества услуг.

Все механизмы TE можно разделить на три группы по политикам QoS:

- Best effort – обработка информации как можно быстрее, но без дополнительных усилий (FIFO, drop tail).
- Мягкий QoS (DiffServ) – сервис с предпочтениями. Приоритетное обслуживание, значения параметров QoS зависят от характеристик трафика.
- Жесткий QoS (IntServ) – гарантированный сервис. Основан на предварительном резервировании ресурсов для каждого потока.

С точки зрения архитектуры сети TE разделяется на средства QoS на узлах, QoS-сигнализацию, решающую вопросы управления сетью, и централизованные политики, обуславливающие заданный уровень сервиса в сети оператора или в сквозном соединении (рис. 35).

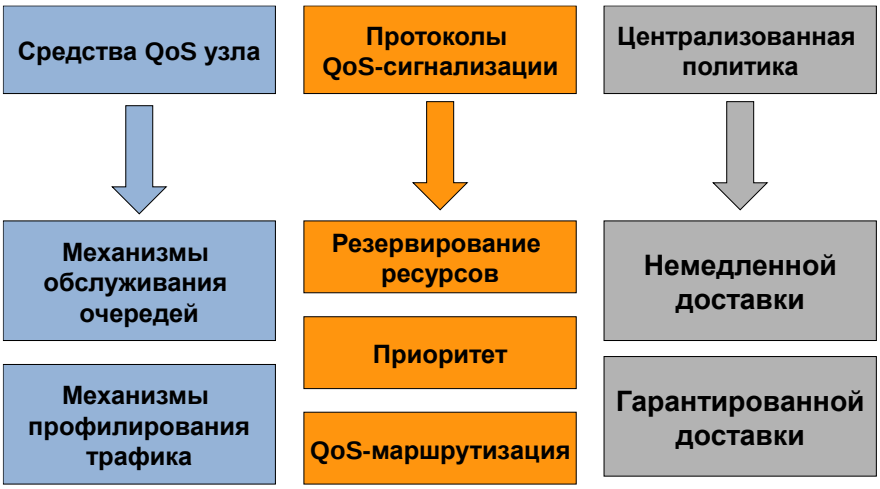


Рис. 35 - Базовая архитектура QoS

Разделение механизмов TE по логическим плоскостям: данных, управления и менеджмента, позволяет определить объект применения данного механизма (рис. 36).

Рис. 36 - Логические плоскости механизмов QoS

Применение механизмов TE возможно на различных сетевых уровнях согласно выбранной оператором стратегии обеспечения QoS.

Транспортный уровень:

- Повторная передача
- Кэширование пакетов на приеме и передаче)
- Подтверждения (квитирование)
- Управление потоком
- Определение тайм-аутов

Сетевой уровень:

– Использование резервирования ресурсов путем организации виртуальных каналов
– Политика обслуживания очередей
– Политика отбрасывания пакетов
– Управление временем жизни пакета
– Маршрутизация

Канальный уровень:

– Управление потоком
– Кэширование
– Повторная передача
– Квитирование

Рассмотрим механизмы QoS согласно архитектуре QoS.

4.1 Средства QoS узла

Управление потоками

Используется на оконечных узлах, реализуется на любом из сетевых уровней. Семейство механизмов управления потоками представлено следующими основными типами:

– Прерывание передачи: при перегрузке передача пакетов источниками трафика прерывается на случайный интервал времени, затем возобновляется с той же интенсивностью.
– Использование динамического окна: размер окна (количество пакетов, посылаемых источником за период) изменяется в зависимости от загрузки буфера.
– Медленный старт: в случае перегрузки источники трафика прекращают передачу, затем посылают пакеты, постепенно увеличивая размер окна (рис. 37). Данный механизм включает в себя оба предыдущих и является самым распространенным в современных пакетных сетях. Это обусловлено прежде всего его применением – он входит в протокол TCP, что делает его одним из мощнейших средств управления сетью.

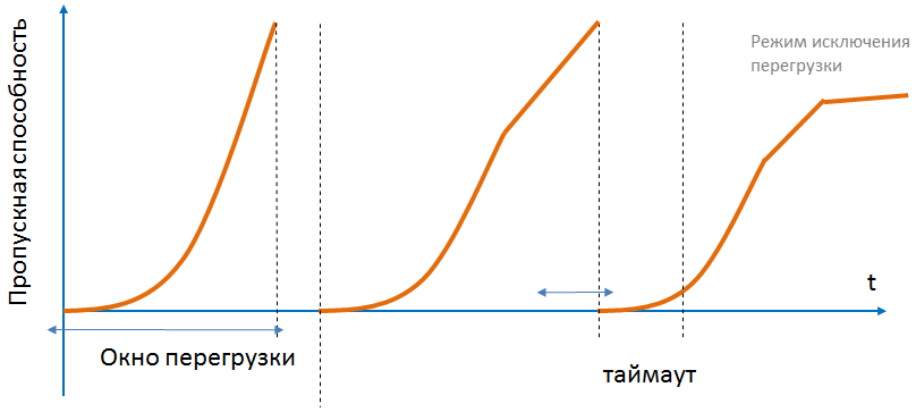


Рис. 37 – Механизм медленного старта

Механизм медленного старта используется для оценки пропускной способности, доступной для потока TCP в данный момент времени. Существует много модификаций TCP, отличающихся нюансами реализации медленного старта, процедурой определения вхождения в медленный старт, механизмами квитиро-

вания. В общих чертах, идея механизма медленного старта состоит в постепенном увеличении динамического окна до момента перегрузки. После этого используется прерывание передачи и возобновление передачи. Однако перед прогнозируемой перегрузкой изменяется скорость увеличения динамического окна передачи, что позволяет сделать более тщательную корректировку скорости передачи TCP-потока. Таким образом, поток передается в режиме исключения перегрузки, так как используется не весь возможный ресурс.

Важным параметром для обнаружения режима исключения перегрузки является round-trip time (RTT) – время ожидания ответа ack, или время двусторонней задержки. Важно отметить, что в IP-сетях в силу их асинхронности измерение времени производится на узле-отправителе, поэтому многие утилиты (например, ping, tracert) оперируют значением времени туда-обратно.

При определении динамического окна корректировка среднего значения RTT происходит по формуле:

$$RTT_m = a \times RTT_m + (1-a) \times RTT_i,$$

где RTT_i - результат i-того измерения, RTT_m – среднее RTT по результатам предыдущих измерений, a - коэффициент сглаживания. Тогда RTO (Retransmission Time Out) время таймаута для повторной передачи может быть определено как:

$$RTO = RTT_m \times b,$$

$$RTT_m = RTT_m + g(RTT_i - RTT_m)$$

$$D = D + d(|RTT_i - RTT_m| - D),$$

$$RTO = RTT_m + 4D,$$

где D - среднее отклонение (для приближенного вычисления, для точного необходимо знать дисперсию), g – коэффициент увеличения таймаута относительно RTT, для коэффициентов задаются значения по умолчанию: a = 0,9; b = 2; D = 0,25; g = 0,125.

С точки зрения QoS данный механизм интересен как средство управления скоростью потока данных. Отметим, что в зависимости от ситуации на сети в момент медленного старта, занимаемый под поток ресурс может существенно отличаться из-за наличия на сети других потоков.

Однако, использование медленного старта может привести к явлению глобальной синхронизации – неэффективного использования полосы пропускания при синхронном вхождении в медленный старт источников трафика в результате одновременных потерь на сети (рис. 38).

Рис. 38 – Причины возникновения глобальной синхронизации

Когда на сети в результате перегрузки происходит потеря пакета, то отсутствие квитанции ack приводит к вхождению узла-отправителя в режим медленного старта. В результате нагрузка на маршрутизатор ослабевает, что уменьшает количество потерь на данном узле. Но в случае, если несколько пакетов от разных потоков, принадлежащих разным узлам, сбрасываются почти одновременно, источники также почти одновременно входят в медленный старт, что может привести к синхронизации. Тогда в сети появляются колебания трафика, обусловленные наложением нескольких процессов увеличения динамического окна (рис. 39). Данный процесс является очень нежелательным, так как приводит к существенному понижению используемой пропускной способности сегмента сети.

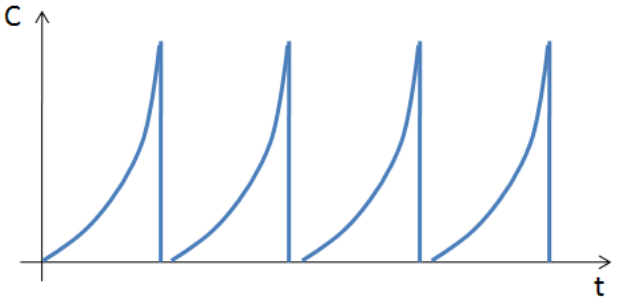


Рис. 39 – Неэффективное использование пропускной способности сети в случае глобальной синхронизации

Для борьбы с данным явлением эффективно использовать методы профилирования трафика.

Методы профилирования трафика

К методам профилирования трафика традиционно относят:

- Drop tail – отбрасывание хвоста: отбрасываются все пакеты, заставшие буфер полным. Используется в best effort.
- RED (Random early detection) – случайное раннее обнаружение: при угрозе перегрузки пакеты из буфера отбрасываются с ненулевой вероятностью.
- Leaky bucket - дырявое ведро: отбрасываются пакеты, не обслужившиеся за установленный период.

– Token Bucket - корзина маркеров (ведро токенов): дозирование трафика с целью уменьшения неравномерности продвижения пакетов.

Рассмотрим каждый из этих методов.

Drop tail является решением по умолчанию. Так как в этом случае отбрасываются все пришедшие и не поместившиеся в буфер пакеты, то именно этот механизм приводит к явлению глобальной синхронизации.

RED (Random early detection) – предотвращает предвзятое обслуживание трафика, эффект глобальной синхронизации, борется с внезапными всплесками трафика, выравнивает джиттер задержки, но приводит к увеличению потерь. Идея механизма заключается в отбрасывании пакетов по какому-нибудь признаку в случае обнаружения перегрузки, т.е. превышения заданного значения минимального размера очереди (рис. 40).

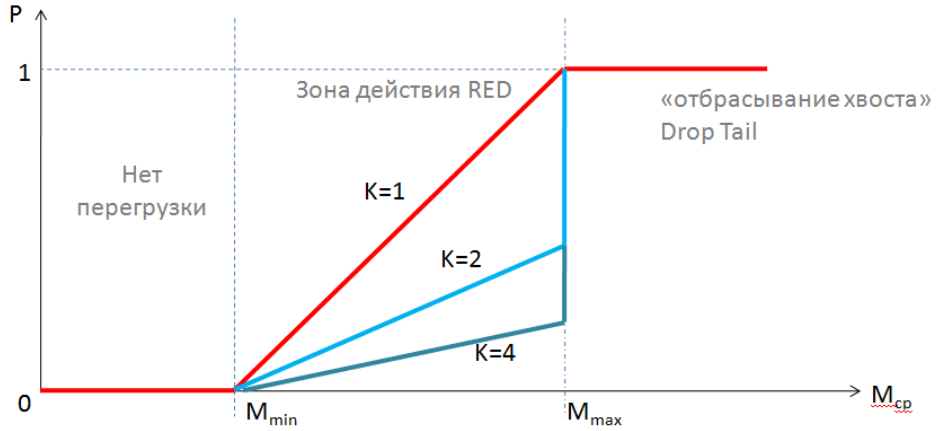


Рис. 40 – Область действия алгоритма RED

На интервале от минимального размера очереди M_{min} до максимального размера очереди M_{max} находится зона действия механизма RED: до M_{min} нет ожидания перегрузки и механизм RED не активен, после превышения M_{max} буфер переполняется и отбрасываются все поступившие на узел пакеты.

Данный механизм содержит два базовых алгоритма:

– алгоритм вычисления среднего размера очереди – для сглаживания трафика и предотвращения реагирования RED на кратковременные всплески трафика (рис. 41). Необходимость усреднения текущего размера очереди обусловлена реагирования механизма RED только на постоянные перегрузки. Усреднение очереди производится как:

$$M_{ср} = M_{ср(t-1)} \times (1 - 0{,}5^n) + M_t \times 0{,}5^n ,$$

где $M_{ср(t-1)}$ – предыдущий средний размер очереди, M_t – текущий размер очереди, n - экспоненциальный весовой коэффициент, по умолчанию $n = 9$.

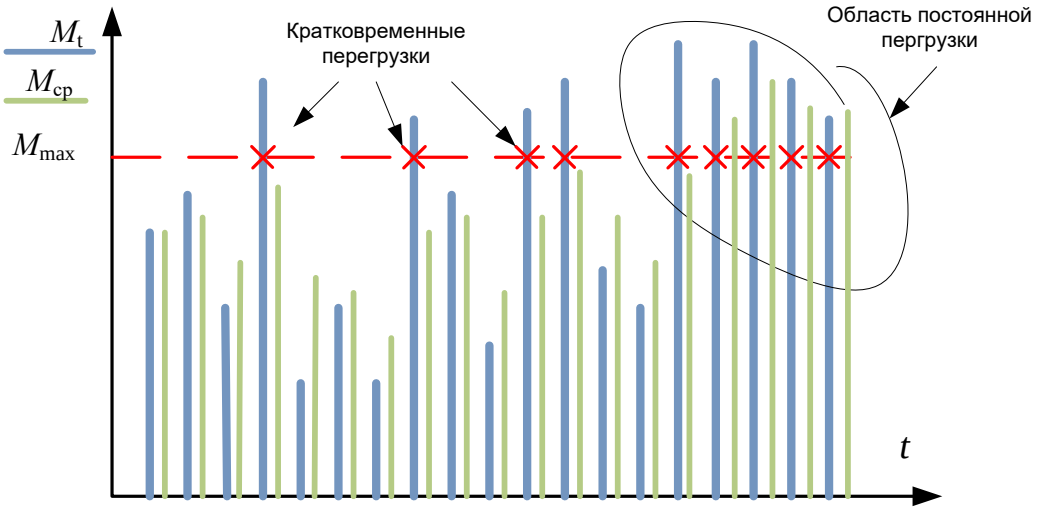


Рис. 41 – Сглаживание текущего размера очереди при n=1

– алгоритм вычисления вероятности отбрасывания пакетов – для определения количества отброшенных пакетов при обнаружении перегрузки.

$$P = \frac{\left(M_{cp} - M_{\min}\right)}{\left(M_{\max} - M_{\min}\right)} \cdot \frac{1}{K},$$

где M_{cp} – средний размер очереди, M_{min} – минимальное пороговое значение среднего размера очереди, M_{max} – максимальное пороговое значение среднего размера очереди, K – знаменатель граничной вероятности (вводится для управления трафиком в зависимости от загруженности узла).

У механизма RED есть несколько модификаций. Одна из самый популярных - Flow WRED (рис. 42), т.е. взвешенное случайное раннее обнаружение перегрузки на основе потока.

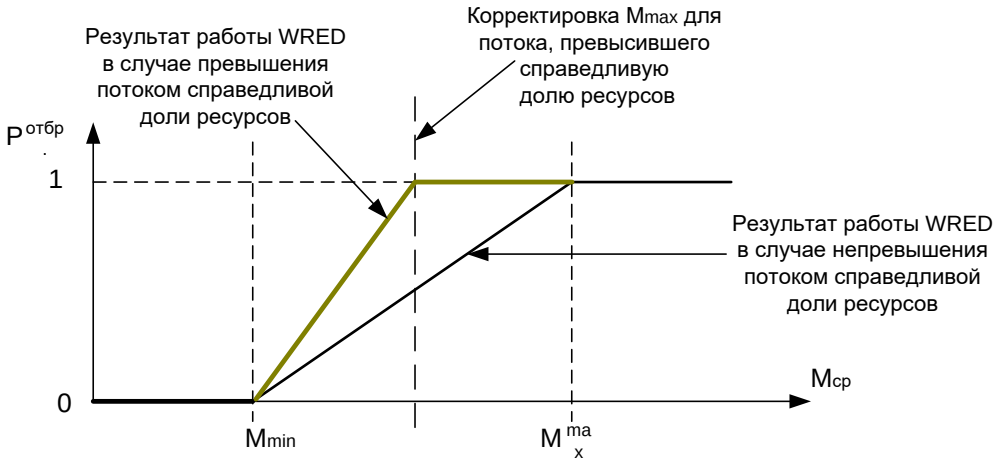


Рис. 42 – Пример работы Flow WRED

Данный механизм классифицирует пакеты в зависимости от приоритета, производит мониторинг состояния активных потоков, корректирует политику отбра-

сывания пакетов путем введения коэффициента масштабирования. Активно применяется для выборочного управления потоками TCP-трафика.

Семейство алгоритмов класса «дырявое ведро» (LB – Leaky Backet) используется практически во всех современных коммутаторах (рис. 43). Идея алгоритма заключается в сглаживании пульсаций трафика. Буфер устройства представляется как ведро с дыркой, через которую со скоростью, согласованной с сетью, т.е. не превышающей скорость обработки пакета обслуживающим прибором, «вытекает» трафик. При этом пакеты, не успевающие обслужиться за период, помечаются специальным битом DE=1 (Discard Eligibility) – признак «окрашивания» пакета. Если пакет с признаком DE=1 не обслужен в течение следующего периода на этом или следующем устройстве, то он должен быть отброшен.

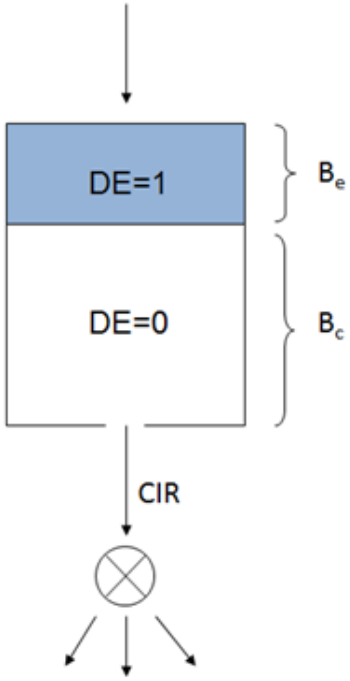


Рис. 43 – Принцип работы алгоритма дырявого ведра

Приняты следующие обозначения:

B_e — допустимое превышение объема пульсации.

B_c — объем пульсации, соответствующий средней скорости CIR и периоду T:

$$B_c = CIR \times T,$$

где CIR – Committed Information Rate: средняя скорость трафика, T — период усреднения скорости.

Token Bucket - корзина маркеров - является механизмом для управления скоростью потока данных на транзитных узлах: дозирования и выравнивания трафика (рис. 44).

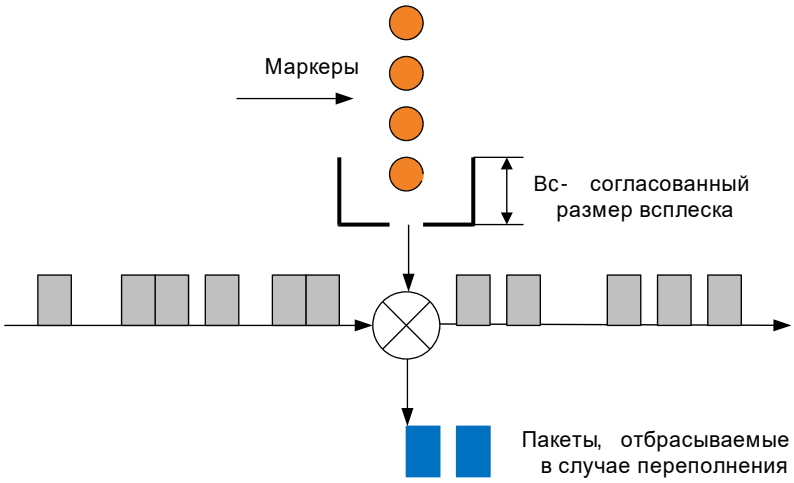


Рис. 44 – Принцип работы корзины маркеров

При поступлении трафика на узел обработка пакетов происходит только после прохождения маркера, за счет чего достигается необходимая скорость потока. Данный механизм имеет две модификации:

– стандартная: не поддерживает резкого увеличения всплеска, допускает потери пакетов (отбрасывание хвоста);
– с возможностью резкого увеличения всплеска: количество маркеров может изменяться при увеличении интенсивности трафика.

Применение корзины маркеров имеет смысл на сегментах сети с агрегацией потоков, так как позволяет уменьшить пачечность общего потока. Отметим, что использование этого механизма может привести к дополнительным потерям.

Механизмы обслуживания очередей

Существующие дисциплины обслуживания можно разделить на бесприоритетные и приоритетные (рис. 45). В современных инфокоммуникационных технологиях успешно применяются оба этих типа. Бесприоритетные циклические дисциплины обслуживания можно встретить в системах балансировки, при опросе беспроводных устройств и в сетях Интернета вещей. Приоритетные дисциплины, особенно с фиксированным приоритетом, активном используются операторами связи для сепарации трафика по категориям пользователей.

Рис. 45 – Классификация дисциплин обслуживания

В буферах транзитных сетевых устройств, таких как коммутаторы и маршрутизаторы, чаще всего реализуются четыре из указанных дисциплин обслуживания и их модификации.

- FIFO (First In First Out) – без использования дополнительных возможностей, используется в best effort.
- PQ (Priority Queuing) – приоритетные очереди, вводится приоритет трафика (высокий, средний, нормальный, низкий).

- CQ (Custom Queuing) – настраиваемые очереди, используется при резервировании ресурсов (до 17 очередей). Вариант: взвешенный алгоритм кругового обслуживания (Weighted Round Robin, WRR).
- WFQ (Weighting Fair Queuing) – взвешенное справедливое обслуживание (рис. 46), позволяет динамически управлять ресурсами (до 256 потоков, ориентируется на ToS). Данный механизм является одним из самых популярных и имеет довольно много модификаций:
  - WFQ на основе вычисления номера пакета
  - WFQ на основе потока
  - CBWFQ на основе класса (до 64 классов)
  - DWFQ – распределенный WFQ
  - DWFQ на основе QoS-группы
  - CBWFQ с приоритетной очередью (LLQ)

В качестве примера рассмотрим WFQ на основе потока (рис. 22). На узел поступают потоки данных различной интенсивности и соответствующие разным типам трафика. Каждому из потоков назначается весовой коэффициент q_i, такой, что

$$\sum_i q_i = 1,$$

где q_i — весовые коэффициенты для i-ой очереди.
Тогда пропускная способность i-го потока:

$$C_i = \frac{q_i \times C}{n},$$

где C — общая пропускная способность узла, n – количество поступающих потоков.

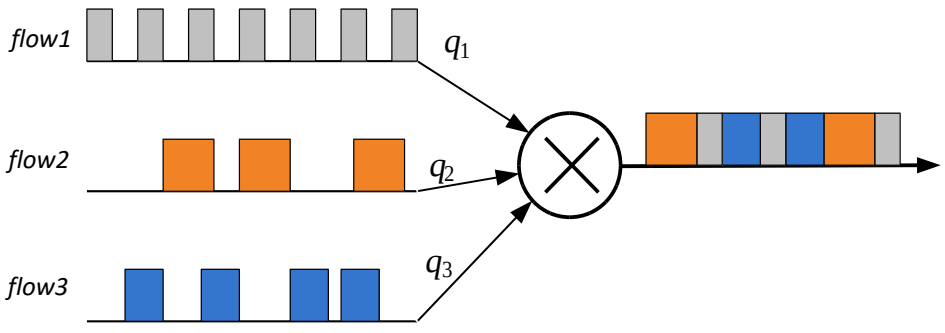


Рис. 46 – WFQ на основе потока

4.2 Средства QoS-сигнализации

Одним из важнейших механизмов QoS-сигнализации является *приоритезация трафика*, за счет чего появляется возможность применения к потокам данных различных политик обслуживания. Маркировать трафик для установления приоритета можно различными методами.

DPI (Deep Packet Inspection) – довольно новая технология глубокого анализа пакетов, которая позволяет идентифицировать и классифицировать трафик, передаваемый по IP-сетям, а том числе вредоносный. В основе данной технологии лежит статистический анализ, использует сигнатурный анализ для классификации трафика по приложениям. Возможно отслеживание и блокирование источников того или иного контента в сети: маркировка цифрового контента позволяет устанавливать источники утечки и распространения нелегальной цифровой продукции.

Методы анализа DPI приведены на рис. 47. Явно заданные правила позволяют приоритезировать трафик по четким признакам со 100% вероятностью. Правила и политики задаются администратором в пределах предоставленных производителем возможностей. Сигнатура — это набор байтов в пакете или файле, позволяющий однозначно определить, к какому приложению, протоколу относится трафик, и классифицировать его. Сканирующий механизм сравнивает анализируемый трафик с известными сигнатурами и при обнаружении соответствия пакет считается классифицированным. Сигнатуры разрабатываются и распространяются производителем оборудования, база сигнатур периодически обновляется. Отметим, что новые протоколы и приложения не всегда правильно классифицируются при устаревших базах. Эвристический анализ – это технология обнаружения трафика по признакам без гарантированной точности. Используется, когда невозможно определить трафик с помощью сигнатурного анализа, результату присваивается значение вероятности определения трафика.

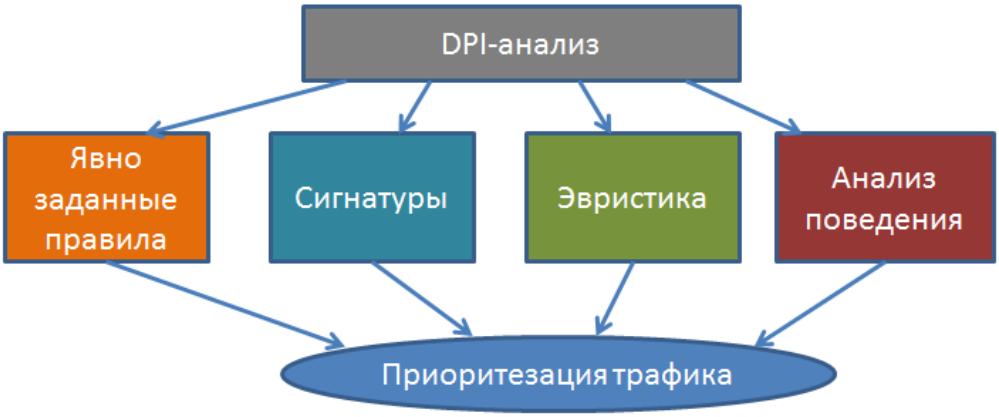


Рис. 47 – Методы анализа DPI

Анализ поведения трафика является одним из старейших методов и активно применяется уже много лет на сетях связи, в том числе ТфОП. В DPI он считается перспективным методом, так как позволяет описать почти любую модель поведения трафика, обладает существенным быстродействием и точностью идентификации. Однако этот метод требует накопления статистики по каждому из типов трафика, для чего существуют базы с моделями поведения трафика.

Идентификация трафика на основе поведенческого анализа может проходить по многим критериям: приложение, генерирующее трафик, перебирает порты; однотипный трафик поступает с множества узлов; трафик имеет определенный размер пакета и т.д.

Идентификаторы управления трафиком для настройки ACL (Access Control List):

- тип протокола.
- VLAN-метки.
- TOS.

Тип протокола может быть идентифицирован по номеру порта. Данная информация извлекается из заголовков протоколов транспортного уровня.

Поле ToS включено в заголовок IP и позволяет уже на сетевом уровне классифицировать трафик.

VLAN-метки задаются на канальном уровне Ethernet согласно рекомендации IEEE 802.1q (рис. 48).

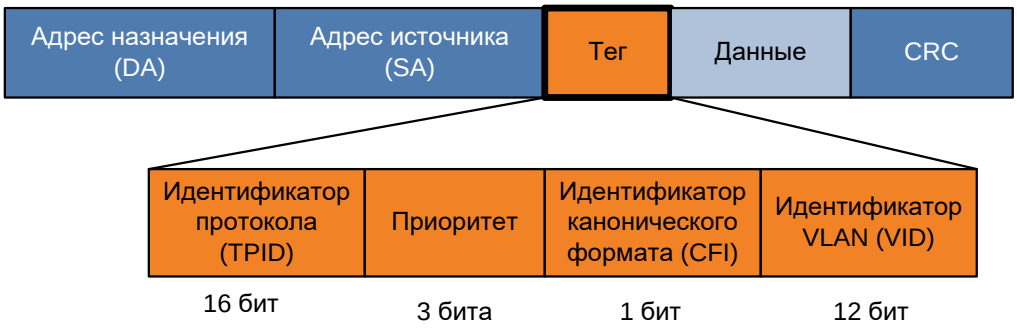


Рис. 48 – Маркированный кадр IEEE 802.1q

Резервирование пути является очень популярным механизмом. Практически резервирование может быть решено как физически применительно к инфраструктуре, так и виртуально. Рассмотрим основные механизмы резервирования пути, которые используются для повышения качества услуг.

RSVP - Resourse Reservation Protocol – протокол резервирования ресурсов. Один из старейших протоколов резервирования, изначально предназначенный для реализации IntServ. Позволяет посылать в сеть информацию о требованиях QoS для каждого потока, работает совместно с IP и протоколами маршрутизации.

Резервирование проводится по адресу получателя (рис. 49). В случает отказа маршрута резервирование происходит заново, таких попыток может быть до 16.

Работает с двумя видами сообщений:

- PATH: запрос на резервирование. Содержит:
  - скорость передачи данных;
  - максимально допустимый размер пульсации трафика.
- RESV: запрос резервирования. Содержит:
  - скорость передачи данных;
  - максимально допустимый размер пульсации трафика.
  - QoS

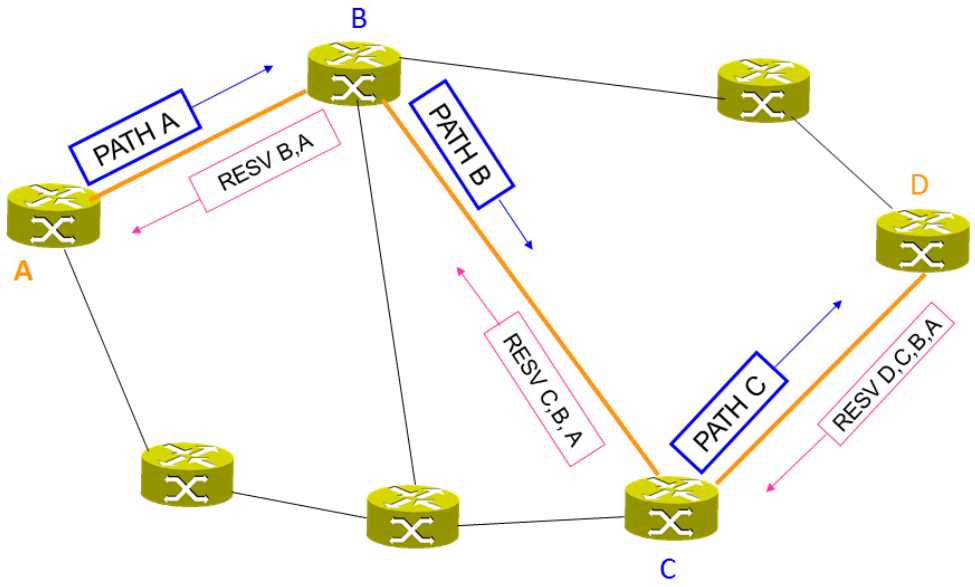


Рис 49 – Организация RSVP-пути

Процесс резервирования пути:

- Узел-отправитель посылает запрос PATH как обычный пакет.
- Каждый маршрутизатор прописывает в своей памяти адрес предыдущего и посылает свой адрес в PATH-запросе.
- Получатель в ответ на PATH генерирует RESV и отправляет по прописанному в PATH пути. Т.о. резервирование происходит в обратном порядке, от получателя к отправителю.
- Маршрутизаторы обрабатывают RESV-запросы, пытаясь предоставить требуемые ресурсы. В случае невозможности предоставления ресурсов резервирование начинается сначала.
- Путь считается установленным, когда отправитель получает RESV. После этого начинается сеанс.

Однако у данного протокола есть ряд недостатков. Важнейший из них – это потеря сетью гибкости при увеличении количества зарезервированных путей. Еще одна особенность – влияние динамической маршрутизации: при резервировании пути в направлении «туда» происходит установление маршрута по узлам, а само резервирование происходит уже в направлении «обратно». Таким образом, может сложиться ситуация, когда при резервировании «обратно» данный путь станет неоптимальным и узел не сможет предоставить запрашиваемый ресурс или уровень обслуживания. В этом случае пакет PATH будет отброшен, и резервирование начнется заново.

Изначально технология MPLS разрабатывалась и как альтернатива RSVP. При сравнении эффективности резервирования MPLS явно выигрывает (рис. 50), хотя также реализована с использованием RSVP.

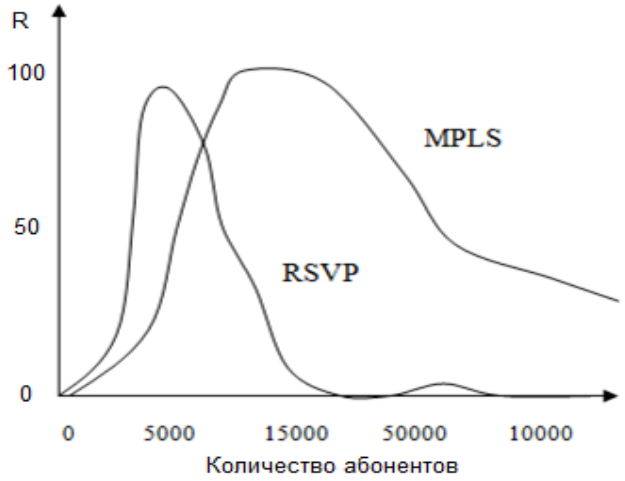


Рис 26 – Влияние количества абонентов на R-фактор (VoIP) при использовании для обеспечения резервирования RSVP и MPLS [10]

Причины перехода от RSVP к MPLS:

- Время сходимости маршрута отличается на три порядка.
- Гибкость и масштабируемость.
- Меньшие вычислительные затраты.
- Повышение безопасности за счет аутентификации и расширений IPSec.

4.3 MPLS

MPLS (Multi-Protocol Label Switching) – коммутация по меткам (RFC 2702, RFC 2283, RFC 2547). Появление MPLS обусловлено необходимостью ускорить процесс обработки пакета на узле за счет замены маршрутизации на коммутацию и отсутствием балансировки нагрузки на маршрутизаторах, т.е. некоторые пути не используются, постоянное переназначение метрик приводит к нестабильности сети, управление трафиком посредством IGP слишком медленное, маршрутизация зависит только от топологии. Пример на рис. 51: используется путь A-C-D-E, путь A-B-D оказывается не загружен.

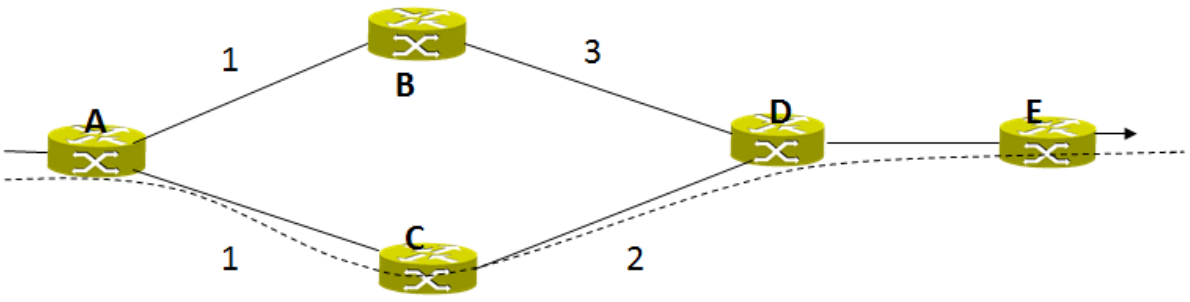


Рис. 27 – Сегмент сети с необходимостью балансировки нагрузки

Цель: ускорить процесс маршрутизации IP-пакетов, расширить возможности обработки трафика в зависимости от типа приложения.

Идея: коммутация меток. Каждый пакет снабжается меткой, которая несет в себе информацию о следующем узле сети. Метка добавляется к пакету (т.е. меж-

ду 2 и 3 уровнем), таким образом, каждый пакет ассоциируется к определенным потоком.

Преимущества: высокая скорость передачи пакетов за счет обработки метки короткого фиксированного размера (20 бит), анализ заголовка IP-пакета только на входе в MPLS-облако, эффективное управление трафиком, поддержка балансировки нагрузки, создание виртуальных каналов.

Рассмотрим пример MPLS-сети (рис 52). Данная сеть снаружи будет выглядеть как два узла: вход с MPLS-облако и выход из него. Существующая внутри такой сети инфраструктура не доступна внешнему наблюдателю, так как работает с системой меток. IP-адреса обрабатываются маршрутизаторами только на входе и выходе из MPLS-сети. LSR (Label Switching Router) – маршрутизатор коммутации по меткам. Последовательность маршрутизаторов (LSR1, LSR2,..., LSRn), через которые проходят пакеты, принадлежащие одному потоку, образует виртуальный путь LSP, коммутируемый по меткам (Label Switching Path).

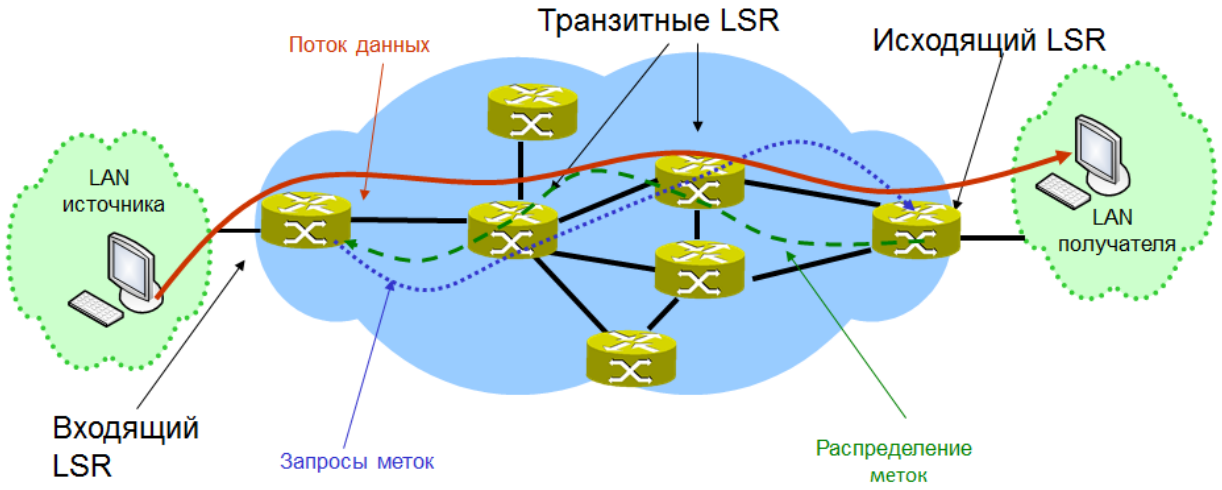


Рис. 52 – Пример MPLS-сети

Метка представляет собой 20-байтную битовую последовательность (рис. 53), содержащую поля, необходимые для поддержки заданного QoS (поле CoS) и агрегирования потоков (признак дня стека меток). В IP-сетях метка вставляется между заголовками канального и сетевого уровней.

Метка: 20 бит

CoS (класс обслуживания): 3 бита

S (признак дна стека меток): 1 бит

TTL: 8 бит

Метка может представлять собой почти любое 20-битное число, кроме специальных от 0 до 15 в десятичной системе:

- зарезервированы:
  - 0: явный ноль IPv4 – пакет должен быть освобожден от метки;
  - 1: метка предупреждения маршрута – пакет должен быть доставлен данному маршрутизатору;
  - 2: явный ноль IPv6 – пакет должен быть освобожден от метки;

– 3: необходимость снятия метки, используется протоколами управления;

- свободны для использования 4-15.

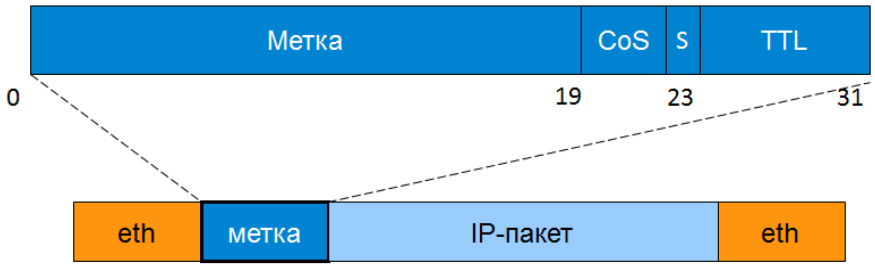


Рис. 53 – Метка MPLS

Существует несколько способов назначения меток:

– Нисходящее назначение: в направлении, обратном к потоку данных. Назначает исходящие метки (рис. 54).
– Нисходящее назначение по требованию: в направлении, обратном к потоку данных. Назначает исходящие метки по требованию предыдущего маршрутизатора.
– Восходящее назначение: по направлению потока данных. Назначает входящие метки.

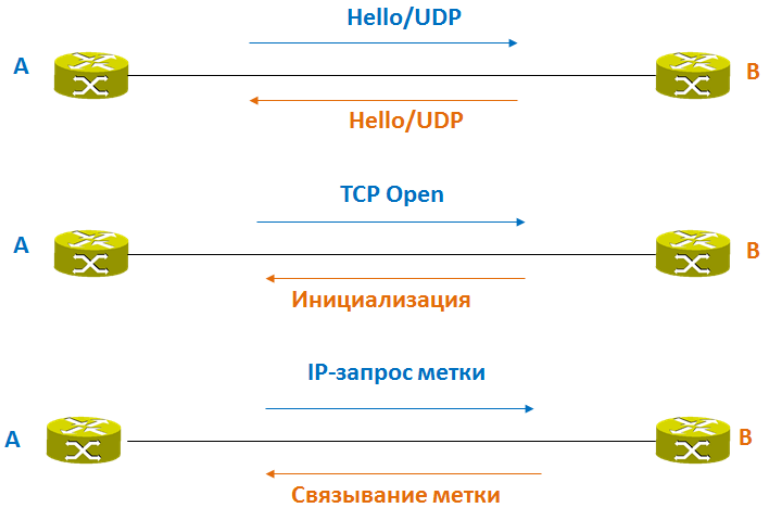


Рис. 54 – Пример процедуры назначения метки

Алгоритм обработки меток (рис. 55):

1. Каждый узел содержит базу меток TIB (Tag Information Base) – т.е. таблицу перенаправлений, содержащую соответствие между полученной и исходящей меткой.

Рис. 55 – Алгоритм обработки меток

2. Узел, получающий пакет, анализирует метку, ищет запись в TIB, изменяет метку на соответствующую и направляет на исходящий порт.
3. Возможен мультикастинг: назначение на одну входящую метку несколько исходящих.
4. Метки назначаются каждым узлом, имеют локальное значения в пределах узла.
5. Если метка не определена: сброс пакета. При этом узел не читает заголовок пакета и не генерирует квитанцию.
6. Пакет может содержать несколько меток – стек (рис. 56). В этом случае решение о коммутации принимается на основе значения последней метки. Используется метод анализа стека меток LIFO.

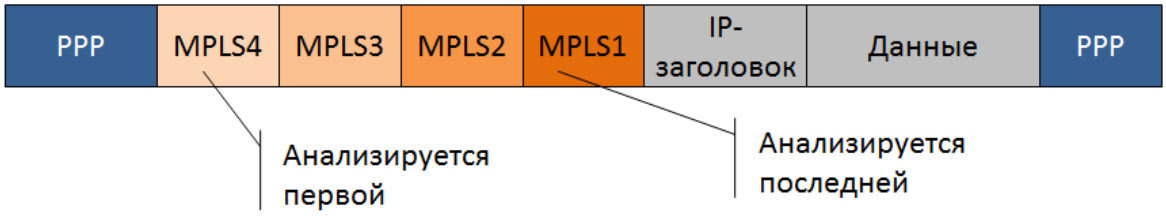


Рис. 56 – Организация стека меток

Входящий узел:

- Получает IP-пакет с адресом получателя, например 192.168.1.5;
- Определяет подсеть 192.168.1.0;
- Добавляет метку к пакету, например, 203;
- Отправляет пакет к следующему узлу

Транзитный узел:

- Получает пакет с меткой, просматривает таблицу коммутации;
- Осуществляет смену меток, например, 203 на 527;

– Передает пакет следующему узлу
– Все остальные транзитные узлы производят аналогичные процедуры.

Предпоследний узел:

– Получает пакет с меткой, просматривает таблицу коммутации;
– Снимает метку (последний узел запрашивает метку 3, т. е. стандартную о последнем LSP-узле маршрута);
– Отправляет пакет к последнему узлу.

Последний узел снимает метку и отправляет IP-пакет получателю.

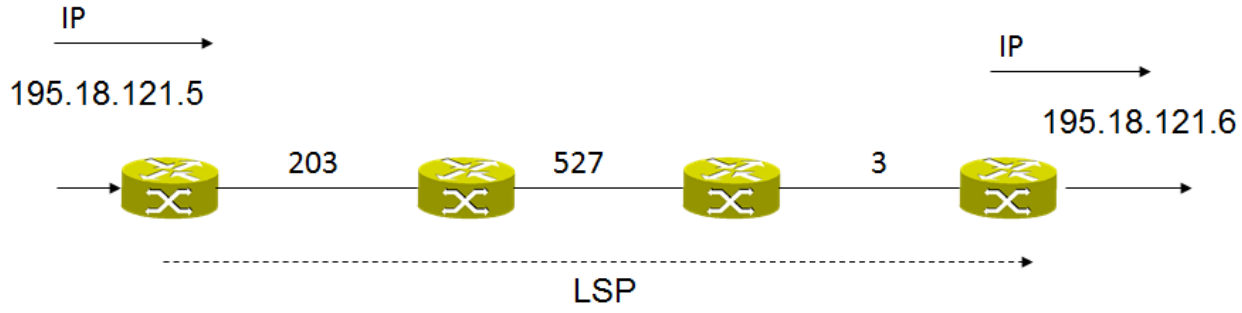


Рис. 57 – Пример обработки LSP

Для организации MPLS-сети с поддержкой QoS используются два протокола (табл. 10).

Таблица 10 – Сравнение LDP и RSVP TE

LDP	**RSVP**
Реализация на сетевом уровне	
Несанкционированные действия: уязвимость сообщений HELLO	Несанкционированные действия: проверка целостности сообщений
Конфиденциальность: нет	Конфиденциальность: аутентификация для каждого из запросов резервирования
Отказ в обслуживании: уязвимость к атакам	Отказ в обслуживании: уязвимость к атакам, проблемы при взаимодействии с IPSec
Нет поддержки резервных маршрутов	Поддержка резервных маршрутов
Время восстановления: больше времени маршрутизации	Время восстановления: минимально за счет резервного туннеля
Обнаружение петель	
Поддержка регулируемой нагрузки: нет	Поддержка регулируемой нагрузки: да
Гарантированная битовая скорость: нет	Гарантированная битовая скорость: да

LDP – Label Distribution Protocol

В функции LDP (RFC 3036) входит: поиск соседей, установление TCP-сессии, обмен метками и ошибками.

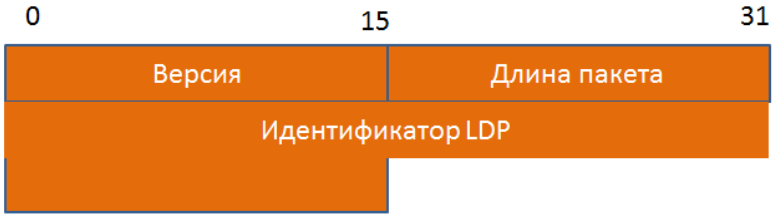


Рис. 58 – Формат пакета LDP

Особенности LDP:

- Поиск соседей происходит по UDP, общение между узлами по TCP.
- Использует доступную на данный момент таблицу маршрутизации. Поддерживает только один маршрут (основной путь). В случае разрушения маршрута восстанавливает резервирование после пересчета таблицы маршрутизации.
- Механизм предотвращения циклов: достигается включением во все сообщения Label Mapping и Label Mapping Request информации о LSR,через которые данные запросы прошли.
- Поддерживает многоадресную рассылку.
- Поддерживает использование сигнатуры TCP MD5.
- Не поддерживает QoS в явном виде: данный протокол позволяет резервировать путь через узлы, но возможности зарезервировать пропускную способность у него нет.

RSVP TE

RSVP TE - Resource ReSerVation for Traffic Engineerin. используется для построения LSP.

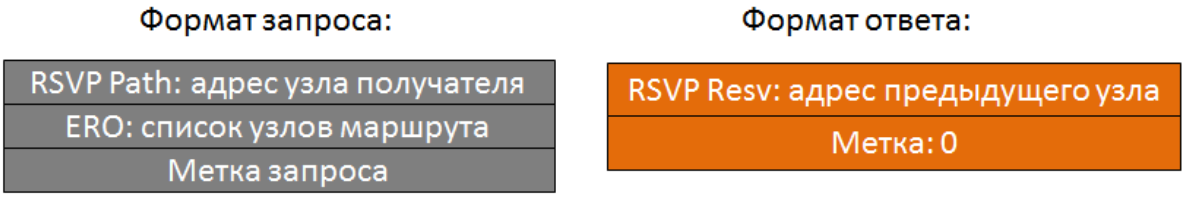


Рис. 59 – Формат пакетов PSVP TE

Возможности:

- Формирование LSP-туннелей с/без требованиями QoS.
- Динамическое изменение маршрутов существующих LSP-туннелей.
- Отслеживание действительного маршрут, проходящего через сформированный LSP-туннель,
- Идентификация и диагностика LSP-туннелей.
- Установление административного контроля сформированного LSP-туннеля.

– Рассылка и объединение запросов выделения меток.
– Не поддерживает классы услуг.
– SESSION, SENDER_TEMPLATE и FILTER_SPEC позволяют однозначно описать туннель.
– Фрагментирует пакеты в случае превышения MTU.

Алгоритм работы:

– Строится TED - Traffic Engineering Database – топология сети с учётом ресурсов.
– Выполняется CSPF - Constrained Shortest Path First – Алгоритм вычисления кратчайшего пути, где: Ingress LSR – входящий маршрутизатор, Egress LSR – конечный маршрутизатор, ERO – Explicit Route Object – список узлов, соответствующих требованиям.

5 МЕТОДЫ БАЛАНСИРОВКИ В IP-СЕТЯХ

Совместно с резервированием путей часто используется балансировка трафика. В настоящее время рассматривают несколько видов балансировок:

- Балансировка трафика – трафик распределяется между альтернативными путями согласно соотношениям пропускной способности. Реализуется с помощью динамической маршрутизации, VPN, наложенных сетей.
- Балансировка кластера – служба балансировки сети распределяет поток данных между несколькими узлами кластера. Реализуется на коммутаторах (с балансировщиком, без балансировщика), с прокси-сервером. Используется в сетях хранения данных, в корпоративных сетях .
- Балансировка DNS – позволяет распределять обращения пользователей к ресурсу между несколькими IP-адресами.

Рассмотрим каждый из видов.

5.1 Балансировка трафика с использованием VPN

Балансировка трафика – набор методов, позволяющих распределять трафик внутри сегмента сети по альтернативным путям для предотвращения перегрузок. Балансировка трафика может осуществляться несколькими методами: традиционной маршрутизации (IS-IS, OSPF), туннелирования (VPN), путем создания наложенной сети.

Механизмы традиционной маршрутизации реализованы в протоколах по состоянию канала. Они позволяют разделять трафик пропорционально метрике и часть направлять по обходным путям. Недостатки: требуют специальной конфигурации протоколов маршрутизации, могут провоцировать нестабильность и образование петель. Достоинства: можно обойтись решениями по умолчанию. Возможности использования протоколов маршрутизации для обеспечения QoS рассматриваются в главе 5.

Методы туннелирования требуют тонкой настройки и подхода в зависимости от требований клиент. VPN – Virtual Private Network – имитируют возможности частной сети в рамках общедоступной, используя существующую инфраструктуру. Особенность VPN состоит в формировании логических связей не зависимо от типа физической среды, таким образом, позволяют обойтись без использования выделенных каналов. Задача туннелирования посредством VPN:

обеспечение в общедоступной сети гарантированного качества обслуживания, а также их защита от возможного несанкционированного доступа или повреждения. Основной принцип при реализации QoS VPN – это возможность управлять полосой пропускания для обеспечения заданного QoS услуг, чтобы гарантировать требуемую полосу пропускания ключевым сервисам. При этом необходимо сохранить критическую пропускную способность для всех сервисов.

Политики QoS VPN

Политики QoS VPN могут поддерживаться как на шлюзах, так и на брендмауэрах (рис. 60) и определяются для следующих категорий:

- Пользователя - ограничивает пропускную способность конкретного пользователя.
- Правил брандмауэра - ограничивает пропускную способность для любого лица, к которому правило брандмауэра применяется.
- Веб-категорий - ограничивает полосу пропускания для URL, выбранного в веб-категории. Для реализации данного ограничения политика назначается через правила брандмауэра.
- Приложений - ограничивает пропускную способность для приложения. Для реализации этого ограничения политика назначается через правила брандмауэра.

При создании VPN можно реализовывать три сценария поддержки политики QoS на брендмауэре: Traffic Policing, Traffic Shaping, Priority Queuing.

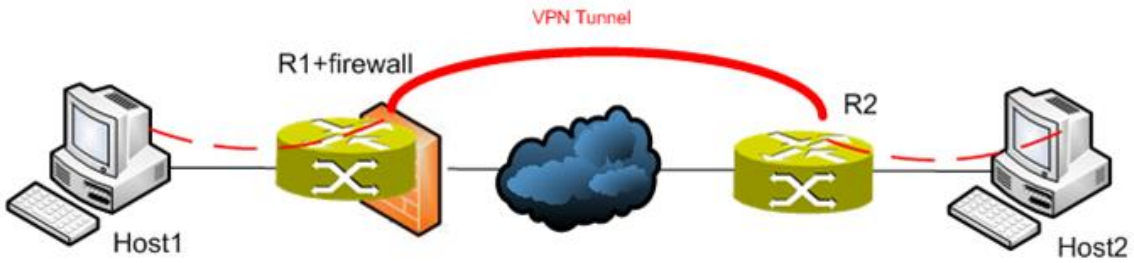


Рис. 60 – Задание политики QoS VPN

Traffic Policing – способ гарантировать настройки политики обслуживания, не допускающие монополизации ресурса, и соблюдение максимальной битовой скорости трафика (рис. 61). При превышении скорости трафика брендмауэр сбрасывает трафик, что приводит к потерям, но минимизирует задержку. Политика обслуживания также задает максимально допустимый всплеск трафика. Для реализации используется корзина маркеров, алгоритм RED.

Traffic Shaping используется для установления соответствия возможностей сетевого устройства и скорости в канале (рис. 62). Данный сценарий позволяет контролировать потери пакетов, джиттер задержки и свободную пропускную способность. Трафик буферизируется и отправляется позже в свободном окне. Traffic Shaping не разделяет типы трафика и не дифференцирует их. Таким образом, механизм вносит задержку, но предотвращает потери, формируя постоянную скорость трафика. Использует алгоритм дырявого ведра.

Рис. 61 – Работа сценария поддержки политики QoS Traffic Policing

Рис. 62 – Работа сценария поддержки политики QoS Traffic Shaping

Priority Queuing позволяет разместить определенный класс трафика в очереди Low Latency Queue (LLQ), которая обрабатывается перед стандартными очередями. Таким образом реализуется политика иерархической поддержки QoS, позволяющая организовать обеспечение QoS пользователей в иерархическом порядке. Этот сценарий используется на брендмауэрах, так как они могут выполнять только LLQ, в отличие от маршрутизаторов, которые могут обеспечить более сложную систему очередей и механизмов QoS: WFQ, CBWFQ и пр. Также этот сценарий используется при приоритезация трафика в Traffic Shaping, когда нет возможности использовать высокоуровневую информацию IP-пакетов.

VPN/MPLS

VPN/MPLS – хорошо масштабируемое решение для балансировки трафика (рис. 63), описано в RFC 2547bis. В данной модели предусматривается два вида LSR:

- P узлы: должны поддерживать маршруты к другим P и PE узлам, а не VPN-маршруты
- PE узлы: поддерживают только непосредственно подсоединенные VPN-маршруты
- VPN могут иметь перекрывающиеся адреса

Основная идея – сделать неуникальные адреса уникальными, заменив группы IP-адресов на RD – Route Distinguisher – признак маршрута. Используется для определения конкретных маршрутов. Это новый тип адреса.

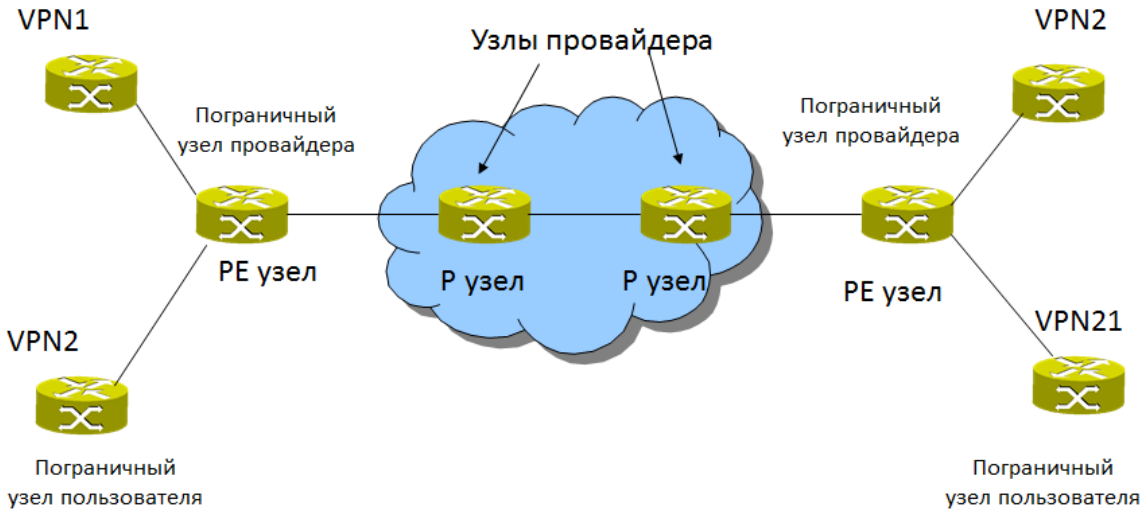


Рис. 63 – Модель взаимодействия с сетью MPLS

Способ: совмещение IP-адреса и некоторого уникального идентификатора. Таким образом, для каждого маршрута в рамках одной VPN будут разные RD.

Комьюнити – сообщества – используются для фильтрации трафика. Обозначаются «цветом». Трансляция комьюнити происходит только в узлах PE. Комьюнити используются только в сети провайдера и только для управления и трансляции.

На рисунке 64 приведен пример адресации VPN/MPLS. Внутренняя адресация должна быть уникальна для провайдера, адресация у клиентов может быть маскарадной.

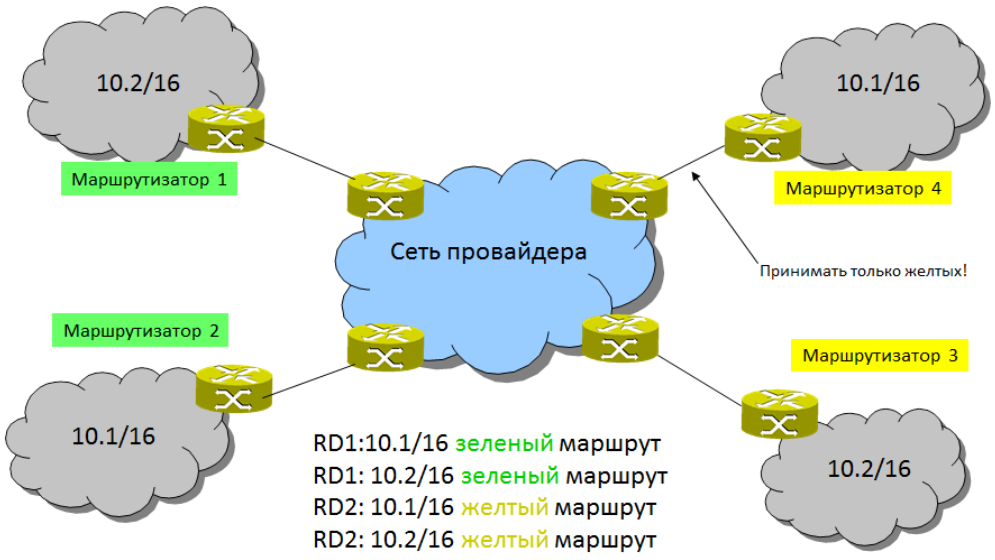


Рис. 64 – Пример адресации VPN/MPLS

При использовании метки VPN/MPLS сначала обрабатывается метка VPN (эта часть маршрута отмечена зеленым), а в MPLS-облаке – метка LSP (рис. 65).

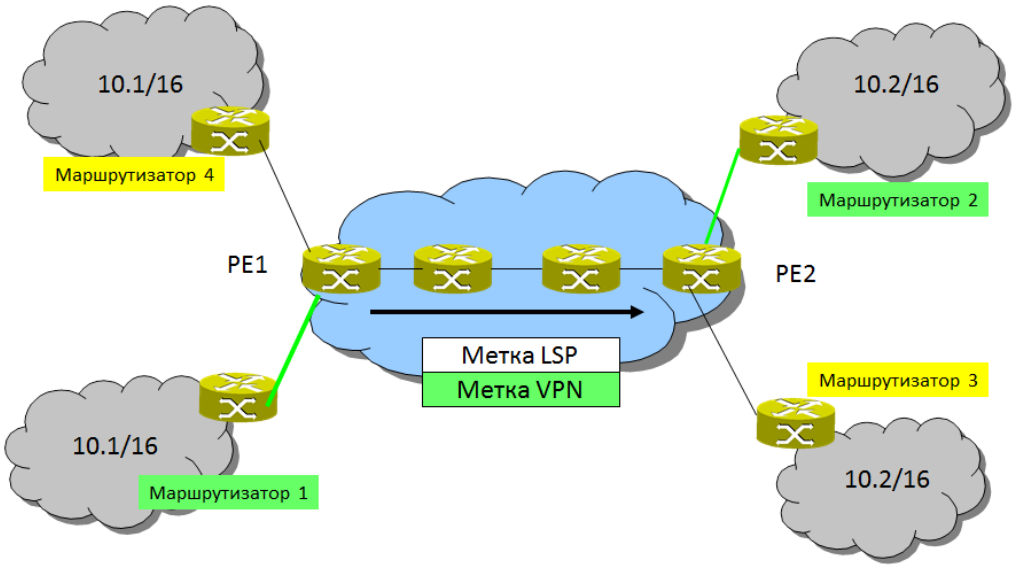


Рис. 65 – Использование метки VPN/MPLS

5.2 Наложенная сеть с функцией балансировки трафика

Еще одним методом балансировки трафика является создание наложенной сети с функцией балансировки трафика или инжиниринг трафика с помощью наложенной сети - Overlay TE (рис. 66). IP-сетям свойственна неравномерная загрузка путей, обусловленная структурой сети и локализацией трафика:

- Топологической особенностью современной глобальной сети является высокий коэффициент кластеризации: вероятность того, что некоторый случайно выбранный узел окажется связанным с географически близким ему узлом.
- Составные маршруты, получаемые на основе метрик нескольких различных протоколов, вообще говоря не оптимальны.
- Для физической топологии Интернет средняя длина пути составляет около 11 проходимых маршрутизаторов с тенденцией к уменьшению.
- Для 30-80% всех маршрутов, проходящих через глобальную сеть, могут быть найдены альтернативные пути, лучшие по показателям пропускной способности, задержек и потерь.
- 20-30% соединений, проходящих через глобальную сеть, постоянно маршрутизируются через перегруженные участки.
- Перегруженные каналы равномерно распределены по всей глобальной сети.
- Вероятности возникновения узкого места внутри сегмента провайдера и в каналах между сетями различных провайдеров приблизительно равны.
- Доля каналов с недостаточной пропускной способностью зависит от положения сетевого сегмента в иерархии, увеличиваясь от 34% у провайдеров Tier-1 до 54% у провайдеров Tier-4.

Реализация системы распределения сетевого трафика на уровне наложенной сети не затрагивает работающих на нижних уровнях традиционных протоколов маршрутизации. С практической точки зрения для реализации наложенной сети

устанавливается специализированное программное обеспечение, которое позволяет объединить некоторые маршрутизаторы в одну систему со своей адресацией и политикой маршрутизации. При этом функционал этих устройств может быть расширен:

- сбор информации о текущем состоянии сети.
- обмен данными о состоянии сети.
- разработка общего плана управления трафиком.

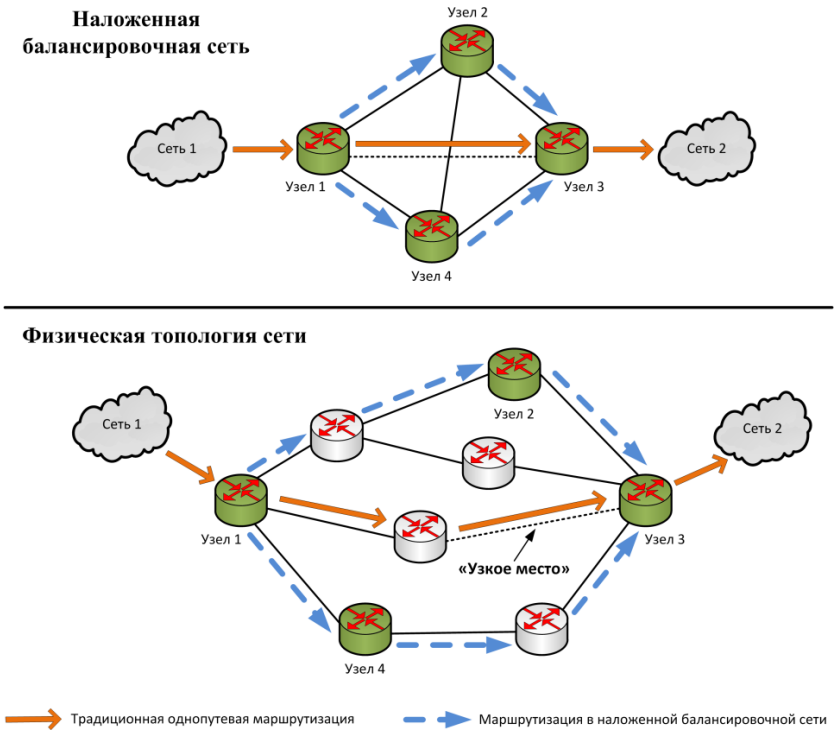


Рис. 66 – Пример создания наложенной сети для балансировки трафика

Существуют системы балансировки трафика с использованием наложенной сети:

- RON (Resilient Overlay Networks) – надстройка над протоколами оптимальной маршрутизации, позволяет находить маршруты в случае отключения BGP. Механизм: система серверов с информацией об активных узлах, инкапсуляция в пакет с маршрутной информацией наложенной сети. Не поддерживает многопутевую маршрутизацию.
- DACoRM (Distributed Adaptive Coordinated Resource management) – использует расширение для протоколов традиционной маршрутизации, позволяет распределять трафик на несколько виртуальных топологий. Время сходимости до 15 минут.

5.3 Балансировка DNS

Еще одним методом балансировки трафика является зеркалирование, которое также называют балансировкой DNS. Зеркалирование ресурсов заключается в присваивании хосту с одним именем нескольких альтернативных ip-адресов, что позволяет распределять трафик посредством традиционной маршрутизации. Различают также Geo DNS – позволяет отдавать разные IP-адреса на основе гео-

графического местоположения клиентов и Bind GeoDNS – аналогичное решение для BIND, использует базу геоданных MadMind.

Рассмотрим пример использования балансировки DNS, для чего воспользуемся утилитой nslookup:

```
C:\Users\admin>nslookup yandex.ru
╤xËтxË:  tellus.sip
Address:  192.168.10.1

Не заслуживающий доверия ответ:
╚ь :     yandex.ru
Addresses:  2a02:6b8:a::a
          77.88.55.66
          77.88.55.55
          5.255.255.5
          5.255.255.55
```

В данном примере при запросе узла с именем yandex.ru возвращен ответ из пяти адресов: адрес IPv6 и 4 адреса IPv4, относящиеся к двум подсетям.

Воспользуемся анализатором трафика и разберем DNS-пакет (рис. 67).

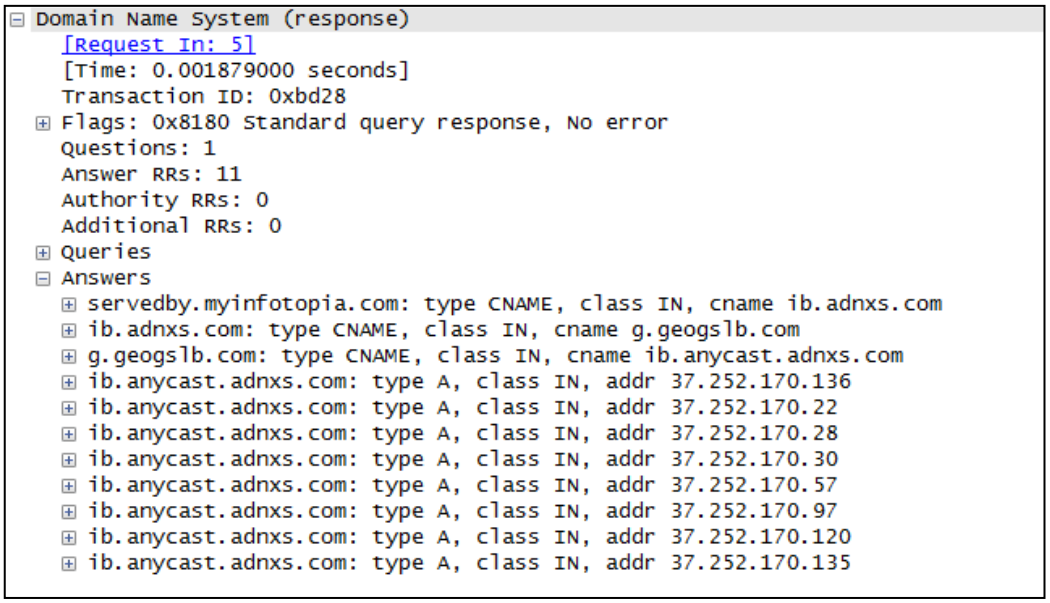


Рис. 67 – Фрагмент пакета ответа DNS

Здесь мы видим использование двух механизмов:

- Редиректор – перенаправление на сайт с дублирующим логическим именем.
- Алгоритм DNS Round Robin (или RR) – создание нескольких DNS-записей класса A и выбор их в циклическом порядке.

Отметим, что при обращении к нагруженным ресурсам использование данного алгоритма крайне важно и внутри кластера. Для распределения входящего потока данных между всеми узлами кластера служба балансировки нагрузки использует широковещательные или многоадресные пакеты протокола второго уровня.

Параметры влияния балансировки нагрузки сети на производительность:

- Накладные расходы ресурсов центрального процессора на узлах кластера (процент вычислительной мощности процессора, необходимый для анализа и фильтрации сетевых пакетов).
- Время отклика на запрос клиента (задержка).
- Пропускная способность к клиентам, увеличивающаяся при росте потока данных от клиентов, который кластер может обработать до достижения максимальной нагрузки на узлы.
- Загрузка коммутатора - не должна критически влиять на пропускную способность портов.

На всех узлах кластера параллельно выполняются одинаковые драйверы службы балансировки нагрузки, объединяемые в единую сеть для обработки входящего потока данных, поступающих на основной IP-адрес кластера. Позволяет заменить маршрутизацию на фильтрацию.

Максимальная пропускная способность узла в составе кластера:

$$RN = C \cdot \left(1 - \frac{N \cdot OF \cdot RN}{R1}\right),$$

где N – количество узлов в кластере, OF — процент ресурсов центрального процессора, используемый службой балансировки нагрузки для фильтрации клиентских запросов при пропускной способности $R1$, С —скорость обслуживания при заданном проценте использования ресурсов центрального процессора, $R1 = C \cdot (1 - OF)$ - максимальная пропускная способность одного узла.

Существует несколько сценариев балансировки внутри кластера:

1. На коммутаторах, с использованием балансировщика – поступающую нагрузку распределяет по серверам специальное устройство. Ответ отправляется не через балансировщик.

1. На коммутаторах, без использования балансировщика – заявка поступает в широковещательном режиме, ответ генерирует первый обработавший заявку узел.

3. На сетевом уровне, с использованием прокси-сервера – данный сценарий похож на первый, только в роли балансировщика выступает прокси и обработка ведется на сетевом уровне.

4. На сетевом и канальном уровне, с использованием двух балансировщиков для повышения отказоустойчивости – в этом случае используется элемент виртуализации для создания возможности резервирования балансировщика в случает перегрузки.

6 ИСПОЛЬЗОВАНИЕ МЕХАНИЗМОВ МАРШРУТИЗАЦИИ ДЛЯ ОБЕСПЕЧЕНИЯ КАЧЕСТВА УСЛУГ

QoS-маршрутизация поддерживается протоколами традиционной динамической маршрутизации (OSPF, IS-IS, IEGRP, BGP). Основывается на маркировке трафика и анализе соответствующих полей в заголовке IP-пакета. Правила пересылки трафика задаются или явно, или по сигнатуре. Основная идея: трафик разного типа направляется по разным маршрутам в соответствии с заданными правилами.

Заметим, что этот класс механизмов в чем-то близок к балансировке трафика, однако существенным отличием является то, что традиционно балансировка не различает трафик по типам, а ориентируется на пропускную способность. При QoS-маршрутизации учитываются требования к показателям QoS в соответствии с типом трафика.

Важным является понятие автономной системы, так как внутри автономной системы и между автономными системами принципы QoS-маршрутизации будут различны.

6.1 QoS-маршрутизация внутри автономной системы

При расчёте маршрута внутри автономной системы необходимо учитывать:

- пропускную способность;
- возможность оборудования обрабатывать поле ToS или DSCP.

Значение поля ToS учитывается при расчете метрики (рис. 68). В RFC 1349 поддерживается 5 типов обслуживания (TOCEC) В настоящее время стандарт устаревший, но оборудование его поддерживает для обратной совместимости.

В RFC 2676 используется расширенное представление ToS: 3 бита зарезервировано, остальные позволяют рассчитывать метрики по задержкам и потерям в виде неявной стоимости. В таблице 11 в качестве примера приведены значения поля ToS для протокола OSPF.

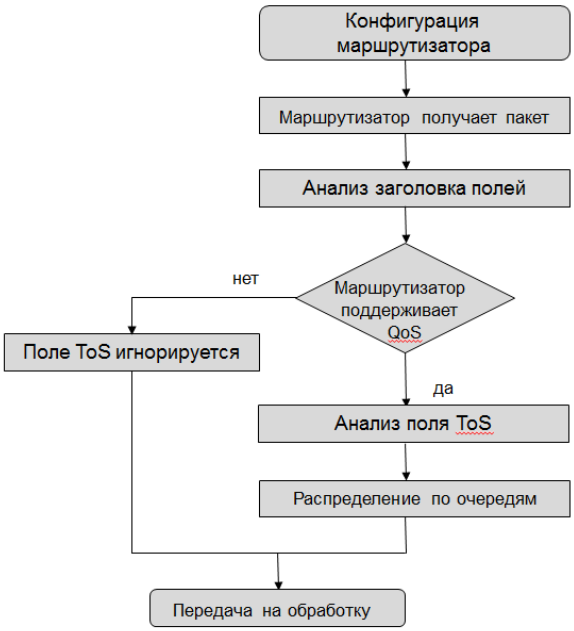


Рис. 68 - Алгоритм обработки пакета при использовании QoS-маршрутизации

Таблица 10 – Значения поля ToS для протокола OSPF согласно RFC 1349 и RFC 2676

Значение поля в десятичной системе		Значение поля в двоичной системе	Требование
RFC 1349	0	0000	Нормальный приоритет
	2	0001	Минимальная стоимость (в деньгах)
	4, 6	0010, 0011	Максимальная надежность
	8, 10, 12, 14	0100, 0101, 0110, 0111	Максимальная пропускная способность
	16, 18, 20, 22, 24, 26, 28, 30	1000, 1001, 1010, 1100, 1011, 1101, 1110, 1111	Минимальная задержка
RFC 2676	32,34,36,38, 40,42,44,46, 48,50,52,54, 58,60,62	10000, 10001, 10010, 10011, 10100, 10101, 10110, 10111, 11000, 11001, 11010, 11011, 11100, 11101, 11110, 11111	Пропускная способность Задержка

В RFC 2676 используется сложная процедура записи метрики в пакете, так как поле для записи значений метрики ограничено 16 битами. Поэтому для расширения возможностей кодирования 16 битами используется экспоненциальное кодирование по основанию 8. Для записи стоимости по пропускной способности 3 бита зарезервированы под основание, 13 на мантиссу, что дает 2^{16}-1 Гбит/с. Для задержки в мкс используется тот же метод, только на основание выделяется 4 бита, на мантиссу 12. Откуда максимальная задержка $(2^{13}\text{-}1)\ x4^{7} \approx 134$ с.

Принцип расчета таблицы маршрутизации:

Пусть матрица KxH, где K - число мест назначения (вершины графа), и H - разрешенное максимальное число переходов на маршруте. Тогда (n; h) маршрут построен во время h-того повторения алгоритма, и состоит из двух параметров:

- bw - максимальная доступная полоса пропускания на маршруте из h переходов между исходным узлом (маршрутизатором) и узлом назначения n;
- сосед – это информация о направлении, связанная с h (или меньше) хопов на пути к узлу назначения n с доступной полосой пропускания bw.

Если S – текущая автономная система, V – промежуточный узел (маршрутизатор), то:

$$bw(S,h) = max\ (\ bw(S,h)\ ;\ min\ (\ bw(V,h)\ ,\ bw(V,S)\ )\ ).$$

Пример графа с QoS-маршрутизацией приведен на рисунке 69. Здесь красным цветом отмечены метрики для трафика данных, синим – для трафика реального времени и черным – потокового.

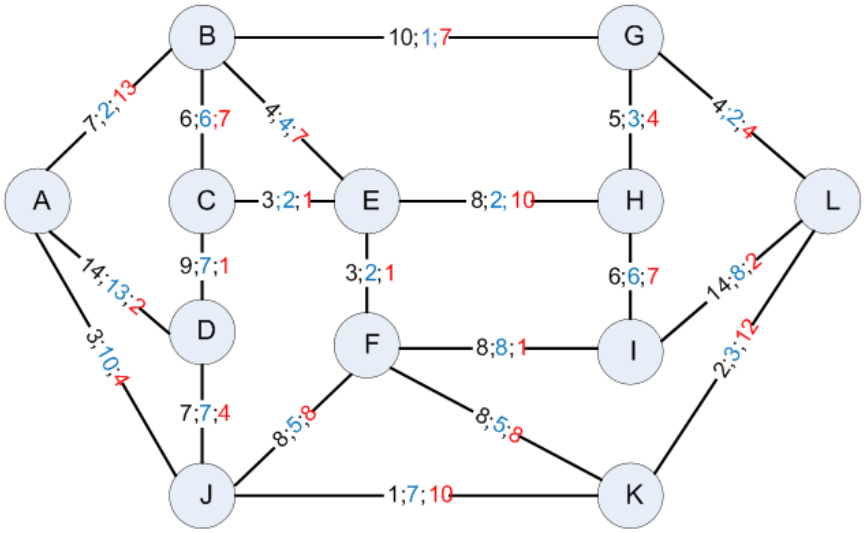


Рис 69 – Пример графа с QoS-маршрутизацией

Требования к QoS-маршрутизации внутри автономной системы:

- Поддержка только маршрутизации от узла к узлу (hop-by-hop).
- Поддержка предварительного расчёта пути.
- Полная интеграция расширений QoS в структуру конфигурационного файла, включая поддержку конфигурации, ошибочную регистрацию, и т.д.
- Разделение пользовательского трафика и трафика управления.
- Взаимодействие с RSVP.
- Политика QoS задается только внутри одной AS.
- Все интерфейсы маршрутизаторов поддерживают QoS.

6.2 Внешняя QoS-маршрутизация

Внешняя QoS-маршрутизация основана на протоколе BGP: маркировке пакетов и маркировке трафика. В BGP подразумевается использование атрибута community. BGP community — атрибут протокола динамической маршрутизации BGP, позволяющий устанавливать определенные метки на передаваемые маршруты. Атрибут позволяет создавать и устанавливать пользовательские значения (установлен только рекомендуемый формат community) и на их основании гибко настраивать фильтры маршрутизатора (рис. 70).

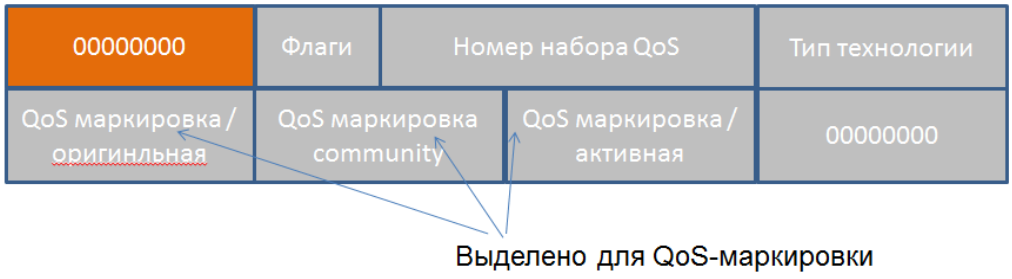


Рис. 71 – Формат записи маркировки с использованием community

Маркировка community содержит 1 октет «Тип», последующие 7 октетов предназначены для QoS-маркировки структуры. «Тип» назначается IANA и отмечает community, передающиеся по AS.

Поле флагов (рис. 72, табл. 12) позволяет гибко управлять маркировками за счет переназначения маркировки или игнорирования ранее присвоенных значений.

Рис. 72 – Формат поля флагов

Таблица 12 – Значения поля флагов

Бит	Флаг	Кодирование
0-1	Не используется	По умолчанию 0
2	P	1 - маркировка сохраняется
3	R	1 - задается перемаркировка
4	I	1- QoS маркировка игнорируется
5	A	1 - задан QoS класс агрегации
6-7	Не используется	По умолчанию 0

Номер набора QoS предназначен для дополнительного упрощения маркировки и позволяет присваивать набор параметров QoS группе потоков. Поле «Тип технологии» отвечает за поддержку технологиями, использующими маркировку (табл. 13).

Таблица 13 – Значение поля «Тип технологии»

Значение	Тип технологии
0x00	DiffServ включен IP (кодирование DSCP)
0x01	Ethernet с помощью 802.1q, используется тэг приоритета
0x02	MPLS с использованием E-LSP
0x03	Кодирование виртуального канала (VC) с использованием отдельных каналов | один канал на класс (ATM VCs, FR VCs, MPLS L-LSPs)
0x04	GMPLS - кодирование временного интервала
0x05	GMPLS - кодирование лямбда
0x06	GMPLS - кодирование волокна

6.3 QPPB

Еще одни методом внешней QoS-маршрутизации является *QoS Policy Propagation Using BGP - QPPB.* QPPB позволяет устанавливать маршрут в таблице маршрутизации с посылаемым классом и приоритетом так, чтобы пакеты, соответствующие маршруту, могли получить связанный QoS. На маршрутизаторах QPPB поддерживается для BGP (IPv4, IPv6, VPN-IPv4, VPN-IPv6), RIP и статических маршрутов. При использовании QPPB маршрут BGP определяется с признаком community, который задает QoS. Отметим, QPPB что имеет совместимости с MPLS.

Маршрутизаторы, которые получают объявление о состоянии канала, вносят маршрут в свою таблицу маршрутизации и устанавливают посылаемый класс и приоритет маршрута в зависимости от признака сообщества.

Для конфигурации QoS-маршрута на IP-интерфейсах используется команда qos-route-lookup, которая поддерживается на интерфейсах типа:

- основные сетевые интерфейсы маршрутизатора;
- интерфейсы группы VPRN (Virtual Private Routed Network);
- интерфейсы SDP (Session Description Protocol).

QPPB может применяться для двух механизмов:

- Координация политики QoS между различными административными областями (AS). AS1 дает знать AS2 об особенностях маркировки трафика, которая должна быть сохранена и в AS2.
- Транспортного дифференцирования в пределах единственной сети или автономной зоны, основанной на особенностях маршрута. Оператор предоставляет дифференцированную услугу в пределах своей сети, трафик должен быть отождествлен с заданным маршрутом.

7 SLA И ПОДДЕРЖКА КАЧЕСТВА СЕРВИСОВ И УСЛУГ

Соглашение об уровне обслуживания (Service Level Agreement - SLA) – это основной документ, регламентирующий взаимоотношения ИТ-сервиса и клиентов, определяет взаимные ответственности провайдера ИТ-сервиса и пользователей этого сервиса. Цель этого документа: дать качественное и количественное описание сервисов, как с точки зрения провайдера, так и сточки зрения клиента. SLA заключается между поставщиком IT-услуг и заказчиком. Например, это может быть соглашение между двумя операторами (Tier-1 и более низким), между IT-подразделением и основным бизнесом компании, между оператором и юридическим лицом.

В SLA фиксируются условия оказания услуг, упорядочиваются взаимоотношения с пользователями, т.е. можно сказать, что SLA – это управление ожиданиями пользователей. Качественным может быть только такое обслуживание, когда каждый, кто подает заявку, знает срок исполнения заявки.

7.1 Модель SLA

В модель SLA входят:

1. Описание сервисов/услуг, предоставляемых в рамках SLA (часть или весь каталог сервисов, предоставляемых ИТ-службой).

2. Описание условий предоставления сервисов / услуг (вплоть до порядка обработки запросов на определенные услуги).
3. Формальные параметры качества предоставляемых сервисов/услуг (время устранения инцидентов; время простоя сервиса целиком и т.п.) и их целевые показатели. Эти параметры качества должны соответствовать бизнес-целям и отражать потребности пользователей в способах оказания услуг.

Таким образом, структура SLA должна содержать как минимум: цель соглашения; указания на субъекты действия соглашения; условия функционирования; установку приоритетов и классов заявок; целевые уровни качества сервиса; процедуру обновления соглашения.

Рассмотрим параметры типовой модели SLA, заключенного между IT-службой и основным бизнесом компании. Заметим, что IT-служба может не являться структурным подразделением компании (например, поддержку сервисов может осуществлять компания - системный интегратор):

- Определение предоставляемого сервиса, стороны, вовлеченные в соглашение, и сроки действия соглашения.
- Дни и часы, когда сервис будет предлагаться, включая тестирование, поддержку и модернизации.
- Число и размещение пользователей и/или оборудования, использующих данный сервис.
- Описание процедуры отчетов о проблемах, включая условия эскалации на следующий уровень. Должно быть включено время подготовки отчета.
- Описание процедуры запросов на изменение. Может включаться ожидаемое время выполнения этой процедуры.
- Спецификации целевых уровней качества сервиса, включая:
- Средняя доступность, выраженная как среднее число сбоев на период предоставления сервиса.
- Минимальная доступность для каждого пользователя.
- Среднее время отклика сервиса.
- Максимальное время отклика для каждого пользователя.
- Средняя пропускная способность.
- Описания расчета приведенных выше метрик и частоты отчетов.
- Описание платежей, связанных с сервисом. Возможно как установление единой цены за весь сервис, так и с разбивкой по уровням сервиса.
- Ответственности клиентов при использовании сервиса (подготовка, поддержка соответствующих конфигураций оборудования, программного обеспечения или изменения только в соответствии с процедурой изменения).
- Процедура разрешения рассогласований, связанных с предоставлением сервиса.
- Процесс улучшения SLA.

На этапе внедрения SLA также оговаривается набор требований:

- Выбор параметров качества, определяющих процесс функционирования сервисов.

- Установленные параметры должны быть достижимы в текущей жизни ИТ-службы.
- Разработка нескольких SLA с разными группами пользователей.
- Разные группы пользователей требуют разных условий оказания однотипного набора услуг.
- Разработка параметров входящей информации, обеспечивающим выполнение ИТ-службой своих обязательств.
- Разработка способов измерения фактических параметров качества.

7.2 Сервисы в SLA

Каталог сервисов позволяет очертить зону ответственности IT-службы, что крайне важно для делегирования полномочий при реагировании на заявки. Внедрение службы технической поддержки (Technical support, Helpdesk, Service desk) связано с определением каталога сервисов и с разработкой SLA.

Время реагирования на заявку должно отличаться для различных групп сервисов, поэтому на этапе регистрации заявки необходимо классифицировать поступившую заявку по типу сервиса, который она затрагивает. Рассмотрим классы сервисов:

Класс 0 – Сервисы, критические для работы большинства пользователей. К ним относятся коммуникационная инфраструктура, проблемы с сервисами этого класса приводят к невозможности взаимодействия.

Класс 1 – Сервисы, критические для работы групп пользователей. К ним относятся все заказные сервисы и подписные, за исключением интернета и электронной почты. Проблемы с сервисами этого класса приводят к невозможности выполнения специальных задач пользователем.

Класс 2 – Сервисы, критические для работы одного пользователя. К ним относятся все базовые сервисы за исключением коммуникационной инфраструктуры, а так же интернет и электронная почта. Проблемы с сервисами этого класса влияют на работу одного пользователя.

Для данных классов сервисов установлены соотвествуюшие целевые уровни качества (табл. 14).

Таблица 14 - Целевые уровни качества сервисов

Уровни качества	**Класс 0**	**Класс 1**	**Класс 2**
Доступность (количество сбоев в год)	Не более 1	Не более 2	Не более 3
Средняя длительность сбоя (часов)	Не более 1	Не более 3	Не более 5
Время отклика	Не заметно пользователю	<1 секунды. Заметно, но не мешает работе	<2 секунд. Заметно, но пользователь может продолжать работу

Необходимо определить и показатели уровней качества сервиса, по которым будет производиться оценка выполнения соглашения. В качестве основных приняты нижеследующие, но в соглашении могут быть оговорены и другие, отвечающие требованиям заказчика:

- доступность сервиса – число сбоев в период обслуживания;
- средняя длительность сбоя сервиса – длительность времени, в течение которого он недоступен потребителю;
- время отклика сервиса – время ожидания, прошедшее после вызова сервиса;
- текущая пропускная способность – пропускная способность канала, оцененная в данный момент. Используется для сравнения со средней пропускной способностью.

7.3 Уровни срочности решения инцидента

Для определения уровней срочности решения инцидентов необходимо установить приоритеты заявок:
Уровень 5 (наивысший) – проблема затрагивает большую часть пользователей или vip-пользователей.
Уровень 4 (высокий) – проблема, затрагивающая группу пользователей.
Уровень 3 (средний) – проблема, затрагивающая одного пользователя.
Уровень 2 (низкий) – проблема, возникающая у пользователей временно.
Уровень 1 (минимальный) – проблема, возникшая единовременно.

Таблица 15 - Уровни срочности решения инцидента

Уровень срочности:	Наивысший уровень срочности	Высокий уровень срочности	Средний уровень срочности	Низкий уровень срочности	Минимальный уровень срочности
Время, выделенное на решение инцидента	4 часа	8 часов	1 рабочий день	5 рабочих дней	10 рабочих дней

Введение четко регламентированного времени реакции для устранения инцидента/предоставления услуги крайне важно и является частью процесса SLM (Service Level Management). Такой подход позволяет также определить эффективность работы службы технической поддержки. К показателям оценки качества работы IT-службы можно отнести количество заявок, поступивших за контрольный интервал времени и процент успешно выполненных заявок; распределение полученных заявок по сервисам, которые они затронули, по классам заявок, типам пользователей, уровням срочности; соотношение инцидентов и запросов на изменения в заявках; среднее время выполнения заявки, полученное для различных классов заявок и уровней их срочности и т.п.

Для технической поддержки используются системы мониторинга, позволяющие осуществлять контроль состояния оборудования и работы сервисов в реальном времени. В корпоративной среде часто используются решения с открытой лицензией, реализованные на основе клиент-сервисной архитектуры и с поддержкой протокола SNMP. На сетях операторов мониторинг сети и выполнение SLA контролируются проприетарными решениями, поставляемыми производителями оборудования.

ЛИТЕРАТУРА

1. Львов Д. С., Глазьев С. Ю. Теоретические и прикладные аспекты управления НТП //Экономика и математические методы. – 1986. – №. 5. – С. 793-804.
2. Малинецкий Г.Г. Проектирование будущего и модернизация России // Препринты ИПМ им. М.В.Келдыша. 2010. № 41. 32 с. URL: http://library.keldysh.ru/preprint.asp?id=2010-41
3. Буйневич М. В. Организационно-техническое обеспечение устойчивости функционирования и безопасности сетей связи общего пользования/Буйневич МВ, Владыко АГ, Доценко СМ, Симонина ОА–СПб.: СПбГУТ,-2013–144с //СПб.: Изд-во СПбГУТ. – 2013.
4. Ibarrola E. et al. QOXPHERE: A new QoS framework for future networks //ITU Kaleidoscope: Building Sustainable Communities (K-2013), 2013 Proceedings of. – IEEE, 2013. – С. 1-7.
5. Ковалгин Ю. А., Вологдин Э. И. Аудиотехника //М: Горячая линия-Телеком-2013.-742 стр. – 2013.
6. Методы оценки качества передачи видео в сетях связи : учебное пособие / М. А. Маколкина - СПб. : СПбГУТ, 2012. - 35 с
7. Симонина О. А., Яновский Г. Г. Характеристики трафика в сетях IP //Труды учебных заведений связи. – 2004. – №. 177. – С. 8-14.
8. Szigeti T. et al. End-to-End QoS Network Design: Quality of Service for Rich-Media & Cloud Networks. – Cisco Press, 2013.
9. Соколов Н. А. Задачи планирования сетей электросвязи //СПб.: Техника связи. – 2012.
10. Гольдштейн Б. С. Технология и протоколы MPLS. – БХВ-Петербург, 2005.
11. Гольдштейн Б. С. и др. Сети связи. – БХВ-Петербург, 2010.
12. Гольдштейн Б.С., Кучерявый А.Е. Сети связи пост-NGN. – СПб: БХВ-Петербург, 2013. 160 с.
13. Иньевски К. Конвергенция мобильных и стационарных сетей следующего поколения. – М.: Техносфера, 2012. 808 с.
14. Кучерявый А. Е., Цуприков А. Л. Сети связи следующего поколения. – 2006.
15. Росляков А. В. и др. Сети следующего поколения NGN //М.: Эко-трендз. – 2008. – Т. 424.
16. Международный Союз Электросвязи. http://www.itu.int/ITU-T/recommendations
17. Степанов С. Н. Основы телетрафика мультисервисных сетей //Изд-во Эко-Трендз. – 2010.
18. Сети связи : учебник для вузов / Б. С. Гольдштейн, Н. А. Соколов, Г. Г. Яновский ; рец.: А. П. Пшеничников, В. В. Лебедянцев. - СПб. : БХВ-Петербург, 2011. - 399 с.
19. Качество обслуживания в сетях связи : научное издание / Ю. Ф. Кожанов ; рец.: Н. А. Соколов, Ю. В. Юркин - СПб. : СПбГУТ, 2014. - 160 с.
20. ITU-T: Key ICT indicators for developed and developing countries and the world (totals and penetration rates). ICT STATISTICS Home Page. http://www.itu.int/en/ITU-D/Statistic
21. Cisco VNI Forecast Highlights. URL: http://www.cisco.com/web/solutions/sp/vni/vni_forecast_highlights/index.html.
22. Дорт-Гольц А.А, Симонина О.А. Механизм управления трафиком посредством балансировки в сетях NGN Информационные телекоммуникационные сети. – 2014. – № 5-6 (93-94). – С.41-45

ПРИЛОЖЕНИЕ

Таблица П1 – Значения параметров E-модели

Параметр	Сокращ.	Ед.	Значение по умолчанию	Реккоменд. значение
Рейтинг громкости передачи	SLR_S	dB	+8	0 to +18
Рейтинг громкости приема	RLR_R	dB	+2	–5 to +14
Рейтинг маскировки местного эффекта	STMR	dB	15	10 to 20
Рейтинг местного эффекта слушающего	LSTR	dB	18	13 to 23
Значение D телефона на передающей стороне	Ds	–	3	–3 to +3
Значение D телефона на приемной стороне	Dr	–	3	–3 to +3
Рейтинг громкости эха говорящего	TELR	dB	65	5 to 65
Взвешенное затухание канала эха	WEPL	dB	110	5 to 110
Средняя задержка канала эха в одном направлении	T	ms	0	0 to 500
Задержка в двух направлениях в 4-проводной замкнутой цепи	Tr	ms	0	0 to 1000
Абсолютная задержка в соединениях, свободных от эха	Ta	ms	0	0 to 500
Число устройств с искажением квантования	qdu	–	1	1 to 14
Коэффициент снижения качества оборудования	Ie	–	0	0 to 40
Коэффициент устойчивости к потере пакетов	Bpl	–	1	1 to 40
Вероятность случайной потери пакетов	Ppl	%	0	0 to 20
Коэффициент всплеска	Nc	dBm0p	–70	–80 to –40
Шум цепи относительно точки 0 дБо	Nfor	dBmp	–64	–
Пороговый шум на стороне приема	Ps	dB(A)	35	35 to 85
Шум помещения на стороне передачи	Pr	dB(A)	35	35 to 85
Коэффициент выигрыша	A	–	0	0 to 20

Симонина Ольга Александровна

КАЧЕСТВО СЕРВИСОВ И УСЛУГ В СЕТЯХ СВЯЗИ

Учебное пособие

Редактор ...

Компьютерная верстка ...

План издания 2016 г., п. …

Подписано к печати 25.11.2016
Объем … усл.-печ. л. Тираж … экз. Заказ …

Редакционно-издательский отдел СПбГУТ
191186 СПб., наб. р. Мойки, 61

Отпечатано в СПбГУТ