

ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА»

Кафедра информационных управляющих систем

Н.Н. Мошак
Модели инфотелетелекоммуникационных транспортных систем

Конспект лекций

для специальности 230201 – Информационные системы и технологии



Санкт-Петербург, 2017

ВВЕДЕНИЕ

Информационное общество XXI века требует создания его технической основы - глобальной информационной инфраструктуры (ГИИ), предоставляющей пользователям необходимые информационные ресурсы и услуги в любое время и из любого места посредством дистанционного доступа, организуемого на основе инфокоммуникационных сетей (ИКС). Это диктуется, с одной стороны, потребностью повседневной жизни людей, а с другой – массовым использованием средств вычислительной техники и новых информационных технологий в структуре хозяйственного, финансового и экономического управления, развитием всемирной электронной коммерции и бизнеса.

Доступ к информационным ресурсам ГИИ реализуется посредством услуг связи нового типа, получивших название услуг Информационного общества, инфоуслуг или инфокоммуникационных услуг. Инфокоммуникационной услугой называется услуга электросвязи, предполагающая автоматизированную обработку, хранение или предоставление информации по запросу с использованием средств вычислительной техники, как на входящем, так и на исходящем конце соединения [ITU-T Recommendation Y.110 (06/98) Global Information Infrastructure principles and framework architecture]. Инфоуслуги - информационные услуги с комплексным предоставлением информации, т.е. предоставление пользователю инфоуслуги в сессии требует организации многокомпонентного потокового соединения с заданными QoS для каждого типа трафика.

К инфокоммуникационным услугам, прежде всего, следует отнести услуги мультимедиа. В соответствии с Рекомендациями ITU-T, услуги мультимедиа делятся на:

- мультимедиа конференции (Multimedia Conference services);
- сбора и накопления информации мультимедиа (Multimedia collection services);
- диалоговые (Conversational services);
- передачи сообщений (Message services);
- выборки информации (Retrieval services);
- распределения (Distribution services) – с индивидуальным управлением процессом предоставления информации со стороны пользователя и без такого управления.

Важным для инфокоммуникационных услуг является понятие «приложение». Приложение определяется как услуга, функциональность которой распределена между оборудованием поставщика услуги и окончательным оборудованием пользователя. Особенность рассматриваемого этапа развития электросвязи заключается в том, что для конечного пользователя создается база для **инфоуслуг приложений**, которые охватывают все виды жизнедеятельности человека (быт, работу, развлечения, медицину, образование торговлю и т. д.). Таким образом, современная бизнес-модель, определяющая участников процесса предоставления инфоуслуг и их взаимоотношения, включает в себя, в отличие от модели традиционных услуг электросвязи, поставщика услуг мультимедиа, который в свою очередь является потребителем услуг переноса, предоставляемых оператором сети связи. При этом основным направлением в развитии услуг электросвязи остается расширение перечня и повышение их качества, а также освоение функций прикладного блока с возможностью переноса мультимедиа в рамках инфокоммуникационной услуги в соответствии с рекомендациями Международного союза электросвязи (МСЭ/ITU) Y-100, Y-101, Y-110, Y-120, Y-130, Y-140 и Y-140.1 [4]. Наблюдаемые в настоящее время высокие темпы роста объемов предоставления инфокоммуникационных услуг позволяют прогнозировать их преобладание на сетях связи в ближайшем будущем.

Инфокоммуникационная сеть строится в соответствии с архитектурой сетей следующего поколения NGN (Next Generation Networks), подчеркивая конвергенцию информационных и телекоммуникационных компонентов в формировании новых сетевых услуг на основе мультисервисной платформы. Согласно определению, приведенному в рекомендациях МСЭ/ITU Y.2001, сеть следующего поколения NGN – это сеть с **коммутацией пакетов**, способная предоставлять телекоммуникационные услуги

посредством широкополосных транспортных технологий, поддерживающих качество обслуживания (QoS), в которой функции услуг не зависят от используемых транспортных технологий. ИКС NGN дает пользователям неограниченный доступ к различным услугам провайдеров и поддерживает обобщенную мобильность, которая позволяет пользователям получить доступ к услугам в любом месте и в любое время.

Область взаимодействия инфокоммуникационной сети в терминах модели взаимодействия открытых систем образует цифровая телекоммуникационная мультисервисная сеть связи (МСС), в которой основные сетевые характеристики интегрального обслуживания трафика различной природы обеспечивает ее коммуникационное мультипротокольное ядро – *инфотелекоммуникационная транспортная система* (ИТС). Инфотелекоммуникационная транспортная система реализует *услуги переноса (bearer service)* мультимедийной информации между сетевыми окончаниями (Network Terminator, NT) без какого-либо анализа или обработки ее содержания. Услуги переноса (доставки) информации характеризуются:

- типами соединений (Connection Type, CT);
- классом качества услуги (Class of Service, CoS);
- параметрами трафика (Traffic Parameters, TP).

Наличие, по крайней мере, двух классов трафика в МСС, требующих различных подходов к их распределению и переносу, приводит к тому, что функциональный или логический профиль любой реализации сетей указанного класса должен включать в себя дополнительную функцию *«совмещения»*, в задачу которой входит «окраска» пакетов типом трафика и создание возможности объединения разнородных информационных потоков различных классов единым образом в общей физической среде.

Кроме того, необходимость предоставления инфокоммуникационной услуги с заданными количественными и качественными параметрами в сессии для переноса мультимедийных объектов в МСС, предопределило необходимость введения в ее функциональный профиль дополнительно к функции *«совмещения»* еще двух функций: 1) функции *«управление резервированием сетевых ресурсов»* для формирования и предъявления заявки на установление мультимедийного соединения и его организации, 2) функции *«контроль допустимости установления соединения»* САС для резервирования затребованных сетевых ресурсов. Это связано со спецификой организации мультимедийного соединения в МСС. Во-первых, для его установления пользователь или соответствующее приложение должны предъявить сети заявку на требуемый физический ресурс с целью обеспечения переноса мультимедийных объектов в сессии с заданными QoS-нормами и, во-вторых, что этот ресурс сеть должна предоставить не нарушив своих обязательств перед уже установленными соединениями.

Необходимость предоставления инфокоммуникационных услуг предполагает новый эволюционный этап интеграции в сетях связи и предъявляет к инфотелекоммуникационным транспортным системам дополнительные требования, связанные со спецификой организации и поддержания мультимедийных соединений (наличие обязательной фазы установления соединений с переменным сетевым ресурсом и заданной многокомпонентной потоковой структурой; блокировки мультимедийных вызовов; появление избыточной мультимедийной нагрузки; фиксации заданного квантиля функции распределения времени пребывания изохронных пакетов в сквозном тракте передачи; смешивания потоков и их синхронизацию в мультимедийной сессии на основе инфокоммуникационной услуги связи; задействование механизмов защиты и др.).

В настоящее время отсутствуют общие принципы и подходы к моделированию и анализу процессов функционирования МСС, базирующиеся на системном подходе к рассматриваемому вопросу, который предполагает, что все элементы системы и все функциональные процессы в ней должны рассматриваться только как одно целое, только в совокупности и только во взаимосвязи друг с другом. Отсутствуют также и математический аппарат для формализованного описания процессов предоставления

инфокоммуникационных услуг связи, имеющий общую методологическую основу построения и позволяющий проводить анализ и синтез всего многообразия реализаций инфотелекоммуникационных транспортных систем.

Суть математического моделирования состоит в замене реальной системы ее образом – математической моделью и дальнейшем изучении модели с помощью реализуемых на ЭВМ вычислительно - логических алгоритмов. Работа с моделью системы дает возможность относительно быстро и без существенных затрат исследовать ее свойства и поведение в различных ситуациях.

Создание национальной МСС является проблемой ближайшей перспективы. В этой связи разработка моделей инфотелекоммуникационных транспортных систем сетей указанного класса является актуальной задачей.

ТЕМА 1. МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ МОДЕЛИРОВАНИЯ СИСТЕМ

1.1. Системный подход к проектированию и моделированию систем

Под *системой* понимают группу взаимосвязанных элементов, действующих совместно с целью выполнения заранее поставленной задачи. Любая система может быть разделена на подсистемы, удовлетворяющие следующим требованиям:

- 1) подсистема является функционально независимой частью системы. Она связана с другими подсистемами, обменивается с ними информацией и энергией;
- 2) для каждой подсистемы могут быть определены функции или свойства, не совпадающие со свойствами всей системы;
- 3) каждая из подсистем может быть подвергнута дальнейшему делению до уровня элементов.

Здесь под элементом понимается подсистема нижнего уровня, дальнейшее деление которой нецелесообразно с позиций решаемой задачи.

Каждая подсистема определяет определенный функциональный профиль системы.

Таким образом, любую систему можно определить в виде набора относительно независимых уровневых функциональных подсистем,

Важнейшие характеристики любой системы в наибольшей степени определяют задачи и методы ее системного, функционального и структурного проектирования.

Подобного рода задачи и методы применительно к проектированию больших систем изучаются *системотехникой*.

Системотехника широко использует достижения других наук, но в ней выработаны и свои принципы, отражающие особенности проектирования больших систем. Основным принципом системотехники является *системный подход*.

Системный подход основывается на свойстве *целостности* проектируемой системы, которое означает:

- наличие у системы единой цели, определяемой ее назначением, достижению которой служат все элементы (части) системы;
- свойства системы как целого не являются простой суммой свойств составляющих ее элементов;
- изменение свойств одного элемента системы вызывает изменение свойств других элементов и системы в целом;
- система взаимодействует с внешним окружением как единое целое.

Инфотелекоммуникационная транспортная система в полной мере обладают свойством целостности. Так, любая ИТС имеет единую цель (назначение) — обслуживание поступающего потока запросов с заданным качеством. Важнейшее свойство ИТС — пропускная способность — зависит от пропускных способностей ее элементов (например, прикладного компонента и менеджера ресурсов (СУБД)), но не является их простой суммой. Более того, пропускная способность, например, прикладного компонента оказывает влияние не только на пропускную способность ИТС в целом, но и на пропускную способность менеджера ресурсов. В процессе обслуживания запросов ИТС взаимодействует с внешним окружением, причем свойства внешнего окружения (характеристики участников соединений) оказывают существенное влияние на свойства ИТС. Так, изменение терпеливости пользователей или среднего времени занятия ими ресурсов прикладного компонента вызывает изменение пропускной способности ИТС в целом. При этом пользователи взаимодействуют с ИТС как с единым целым, т. е. для них не имеет значения внутренняя структура ИТС.

Системный подход подчеркивает значение комплексного учета и оценки как можно большего числа внешних и внутренних факторов, которые могут влиять на характеристики моделируемой системы, с тем, чтобы не упустить действительно существенные факторы. Таким образом, системный подход к проектированию ИТС заключается в комплексном учете и оценке как можно большего числа факторов, характеризующих внешнее окружение, в оптимизации построения и функционирования ИТС в целом с учетом взаимосвязи и взаимозависимости этих факторов.

В дальнейшем, при построении моделей ИТС мы будем использовать **системный подход**.

1.2. Основные процедуры проектирования систем

Системный подход дает общие принципы, которыми необходимо руководствоваться при постановке и решении задач проектирования ИТС на различных этапах. Однако он не раскрывает с помощью каких процедур и в какой последовательности применяемых процедур достигается решение этих задач.

Процедуры проектирования определяют совокупность методов, правил, приемов, применяемых в определенной последовательности для решения задач проектирования. Основными процедурами проектирования ИТС являются процедуры *синтеза и анализа*.

С помощью *процедуры синтеза осуществляется выбор варианта построения системы*, реализующего заданный набор функций в заданном базисе элементов и обеспечивающего требуемые характеристики системы. Обычно при проектировании ставится задача выбора наилучшего (оптимального) в смысле заданного критерия варианта построения системы. В этом случае говорят о *процедуре оптимального синтеза*.

Процедура оптимального синтеза применима для решения задач проектирования различных систем на различных этапах, так как эти задачи отличаются одна от другой только заданным набором реализуемых той или иной системой функций и базисом элементов и с формальной точки зрения аналогичны.

Выбор оптимального варианта построения системы основан на оценке значений характеристик и критерия для всех рассматриваемых вариантов. Выполнение такой оценки для заданного варианта построения системы осуществляется с помощью процедуры анализа. Таким образом, процедура синтеза обязательно включает в себя *процедуру анализа*.

1.3. Формулирование задачи оптимального проектирования системы

Процедура синтеза начинается с формулирования задачи проектирования, решаемой на том или ином этапе (рис. 1.1). Формулирование задачи проектирования заключается в разработке множества альтернатив (вариантов) построения проектируемой системы, определении совокупности ограничений на значения характеристик и параметров системы для каждой альтернативы и выборе критерия оптимальности для сравнения различных альтернатив между собой.

Пусть A — множество альтернатив построения системы. Каждой альтернативе $A_i \in A$ можно сопоставить вектор независимых параметров $\vec{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_{n_i}^{(i)})$, отражающий в количественном виде некоторые первичные свойства системы (состав функций, алгоритм функционирования, количество комплектов различных типов, количество управляющих устройств, их быстродействие и др.). Значения каждого параметра $x_k^{(i)}$ могут выбираться из определенного для альтернативы A_i множества $X_k^{(i)}, k = \overline{1, n_i}$.

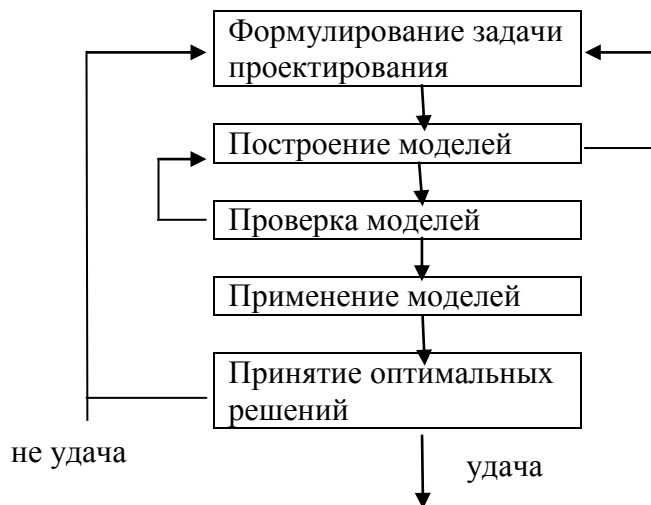


Рис. 1.1 – Алгоритм решения задачи проектирования

Декартово (прямое) произведение этих множеств задает область допустимых значений (граничные условия) для вектора параметров $\vec{x}^{(i)} \in X^{(i)} = X_1^{(i)} \times X_2^{(i)} \times \dots \times X_{n_i}^{(i)}$.

Характеристики системы y_k отражают в количественном виде некоторые производные от первичных или вторичные свойства системы (например, вероятность потери мультимедийного вызова, среднее время реакции системы на запрос, стоимость системы и др.), т. е., $y_k = f_k^{(i)}(\vec{x}^{(i)}, k = \overline{1, m})$. Набор значений характеристик системы - образует вектор характеристик $y = (y_1, y_2, \dots, y_m)$, областью допустимых

значений которого является множество $Y = (Y_1 \times Y_2 \times \dots \times Y_m)$, где $Y_k, k = \overline{1, m}$ — множество допустимых значений (функциональные ограничения) для характеристики y_k .

В качестве критерия оптимальности K выбирается обычно интегральная характеристика системы, отражающая совокупность важнейших вторичных свойств системы, т. е. $K = \varphi(\bar{y})$. Поскольку $y_k = f_k^{(i)}(\bar{x}^{(i)}, k = \overline{1, m})$, критерий оптимальности также является функцией параметров системы: $K = \Phi^{(i)}(\bar{x}^{(i)})$.

В зависимости от выбранного критерия задача оптимального проектирования системы может быть теперь формализована в виде:

1) найти такие $A_{i \max} \in A$ и $\bar{x}_{\max}^{(i)} \in X^{(i)}$, что

$$K_{\max} = \Phi^{(i)}(\bar{x}_{\max}^{(i)}) = \max_{A_i \in A} \max_{\bar{x}^{(i)} \in X^{(i)}} \Phi^{(i)}(\bar{x}^{(i)}), \text{ при } \bar{y} \in Y \dots\dots\dots (1.1)$$

2) найти такие $A_{i \min} \in A$ и $\bar{x}_{\min}^{(i)} \in X^{(i)}$, что

$$K_{\min} = \Phi^{(i)}(\bar{x}_{\min}^{(i)}) = \min_{A_i \in A} \min_{\bar{x}^{(i)} \in X^{(i)}} \Phi^{(i)}(\bar{x}^{(i)}), \text{ при } \bar{y} \in Y \dots\dots\dots (1.2)$$

Из соотношений (1.1) и (1.2) видно, что задачу оптимального проектирования системы можно разделить на две самостоятельные задачи: выбора оптимального вектора параметров для заданной альтернативы построения системы и выбора оптимальной альтернативы из заданного множества при фиксированном для каждой альтернативы векторе параметров. Эти задачи в теории проектирования систем получили название задач *параметрического и соответственно структурного синтеза*.

Необходимо отметить, что формулирование задачи оптимального проектирования в окончательном виде (1.1) или (1.2) осуществляется итеративным путем по мере накопления знаний о построении и работе проектируемой системы.

1.4. Процедура принятия оптимального решения

Процедура принятия оптимального решения в соответствии с формализацией задачи оптимального проектирования системы в виде (1.1) или (1.2) сводится к решению задач параметрического и структурного синтеза системы.

Задачу параметрического синтеза для альтернативы $A_i \in A$ при оптимизации по максимуму критерия можно формализовать в следующем виде:

найти вектор параметров $\bar{x}_{\max}^{(i)} \in X^{(i)}$, такой, что

$$K_{\max} = \Phi^{(i)}(\bar{x}_{\max}^{(i)}) = \max_{\bar{x}^{(i)} \in X^{(i)}} \Phi^{(i)}(\bar{x}^{(i)}), \tag{1.3}$$

при $\bar{y} \in Y$.

Как правило, аналитические зависимости для функций $\Phi^{(i)}$ и $f_k^{(i)}, k = \overline{1, m}$ (если их удастся получить), настолько сложны, что решение этой задачи известными аналитическими методами (например, методом множителей Лагранжа) является практически невозможным. Поэтому наиболее широко для решения задачи параметрического синтеза применяются численные методы, использующие аналитические и численные модели системы, а также эвристические методы.

Численные методы основаны на вычислении оптимального значения функций $\Phi^{(i)}$ и $f_k^{(i)}$ в соответствии с математически обоснованным алгоритмом, гарантирующим

при определенных условиях нахождение этого оптимума. Для решения задачи параметрического синтеза в зависимости от вида функций $\Phi^{(i)}$ и $f_k^{(i)}$ могут применяться методы линейного (при линейных функциях), динамического (при аддитивной функции $\Phi^{(i)}$) нелинейного (при нелинейных функциях) и, в частности, дискретного (если параметры принимают значения из дискретного множества) программирования, а также методы регулярного и случайного поиска.

Эвристические методы основаны на вычислении оптимального значения функции в соответствии с алгоритмом, обобщающим и формализующим опыт решения аналогичных задач оптимизации человеком. Эвристические методы не имеют математического обоснования и поэтому не гарантируют получения строго оптимального решения, однако их применение дает в ряде случаев решение, достаточно близкое к оптимальному (субоптимальное).

После решения задач параметрического синтеза для всего множества альтернатив решается задача структурного синтеза системы, которая теперь может быть формализована в виде:

$$\text{найти альтернативу } A_{i_{\max}} \in A \text{ такую, что}$$

$$K_{\max} = \Phi^{(i)}(\bar{x}_{\max}^{(i)}) = \max_{A_i \in A} \{K_{\max}^{(i)}, i : A_i \in A\}, \quad (1.4)$$

при $\bar{y} \in Y$.

На практике число различных альтернатив построения системы обычно невелико, поэтому задача структурного синтеза решается методом последовательного перебора альтернатив (сравнения чисел $K_{\max}^{(i)}$).

1.5. Классификация моделей системы

Задачи структурного и параметрического синтеза системы, в сформулированном выше виде могут быть решены, если известны функциональные зависимости критерия оптимальности и характеристик системы от вектора параметров для каждой альтернативы. Для выявления этих функциональных зависимостей используется процедура построения *моделей системы*.

Моделирование является одним из наиболее важных методов научного исследования и экспериментирования. Модель представляет собой абстрактную или физическую систему, адекватно отображающую или воспроизводящую свойства реальной системы — объекта моделирования. Широкое применение моделей в процессе проектирования систем основано на их способности при определенных условиях замещать объект моделирования так, что изучение модели дает необходимую информацию об этом объекте. Подобная способность модели обеспечивается той или иной степенью адекватности (соответствия, подобия) между моделью и реальной системой.

По степени адекватности различают изоморфные и гомоморфные модели (рис. 1.2). *Модель называется изоморфной системе*, если существует взаимно-однозначное соответствие между элементами модели и системы и между отношениями (связями, зависимостями) элементов в модели и в системе. В процессе проектирования изоморфные модели находят ограниченное применение, так как они оказываются такими же сложными, как и сама система, и не упрощают процедуру анализа ее характеристик.

Однако изоморфные модели позволяют получить в принципе точные оценки характеристик системы. Более удобными являются *гомоморфные модели*, отображающие или воспроизводящие только существенные с точки зрения решаемой задачи анализа свойства (элементы и отношения между ними) реальной системы.

Благодаря этому процедура анализа характеристик проектируемой системы значительно упрощается, однако результаты анализа носят, естественно, приближенный характер.

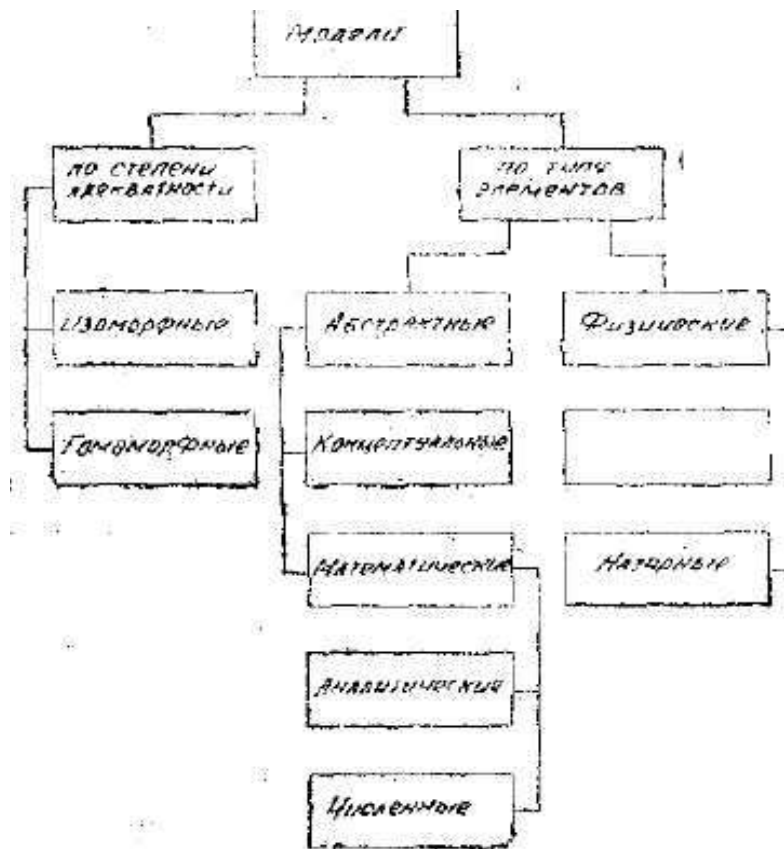


Рис.1.2 -
моделей

Классификация

По типу элементов различают *абстрактные (знаковые)* и *физические (предметные)* модели.

Абстрактная модель представляет описание проектируемой системы на некотором абстрактном языке. Элементами такой модели служат абстрактные понятия и знаки — формулы, схемы, чертежи, графики, таблицы, предложения в некотором алфавите естественного или искусственного языка.

Простейшим видом абстрактных моделей являются **концептуальные модели**. Концептуальная модель выявляет причинно-следственные связи между характеристиками (критерием оптимальности) системы и ее параметрами. Основное ее назначение — выявление набора параметров, учет которых необходим для оценки той или иной характеристики системы. Одна и та же система может быть представлена различными концептуальными моделями в зависимости от цели исследования (вида характеристики).

Важнейшим видом абстрактных моделей являются **математические модели**, дающие описание системы на некотором математическом языке, т. е. в терминах используемой математической теории.

Аналитическая модель выявляет количественные отношения между характеристиками (критерием) и параметрами системы в форме функциональных зависимостей $\bar{y} = F^{(i)}(\bar{x}^{(i)})$ или $K = \Phi^{(i)}(\bar{x}^{(i)})$. Аналитическая модель позволяет провести анализ соответствующей характеристики системы в наиболее общем виде, однако ее построение требует существенного упрощения анализируемой системы.

Для аналитических моделей функционирование элементов системы записывается в виде некоторых функциональных соотношений (алгебраических, интегро-дифференциальных уравнений и т.п.) или логических условий. Построение аналитических моделей сети основано на применении теории массового обслуживания. При построении аналитических моделей сети используются Марковские и полумарковские случайные процессы, вложенные цепи Маркова, диффузионные процессы и т.д. В последнее время в качестве моделей сетей связи широко применяются их модели в виде сетей массового обслуживания (СеМО). В основе этого подхода, как правило, лежат математическая теория массового обслуживания и теорема Джексона, определяющая условия, при выполнении которых сеть можно рассматривать как совокупность независимых систем массового обслуживания.

Аналитические модели могут исследоваться тремя способами:

- *аналитическими методами*, когда стремятся получить в общем виде явные зависимости для искомых величин,
- *численными методами*, когда из-за невозможности решить уравнения в общем виде, есть возможность получить числовые результаты при некоторых начальных данных,
- *качественными методами*, когда не имея решения в общем виде, можно найти некоторые свойства решения, например, оценить устойчивость решения и т.п..

Аналитические методы позволяют провести наиболее полные и исчерпывающие исследования моделируемой системы, а также применять методы оптимизации. Однако это удастся лишь для сравнительно простых систем. Применение аналитических методов для более сложных систем связано с большей по сравнению с другими методами степенью абстрагирования и упрощения реальности. Поэтому аналитические методы исследования используются обычно для первоначальной оценки характеристик всей сети или отдельных ее подсистем. Ограниченные возможности аналитических методов способствовали развитию численных методов.

Численные методы по сравнению с аналитическими применимы к более широкому классу систем, однако, получаемые при этом решения зависят от исходных данных. Одним из наиболее перспективных методов численного анализа сетей связи является метод, основанный на создании алгоритмов автоматизированного составления уравнений Колмогорова для систем, описываемых Марковскими процессами, и их решения на ЭВМ.

Качественные методы исследования систем позволяют определять такие свойства системы как непрерывность и устойчивость.

Структура и характер процесса функционирования проектируемой системы, подлежащие отображению в модели, могут быть столь сложными, что построение аналитической модели, описывающей в явном виде функциональные зависимости характеристик системы от ее параметров, становится невозможным. В этом случае используются *численные модели*, описывающие 1) либо последовательность действий, которые должны быть выполнены над численными значениями параметров для получения численных значений характеристик, 2) либо структуру и процесс функционирования системы во времени в логико-математической форме на некотором алгоритмическом языке.

Численные модели последнего типа описывают, в сущности, процесс имитации функционирования реальной системы на ЭВМ и называются *имитационными*. Имитационная модель – это набор алгоритмов, генерирующий некую последовательность, которая по своим характеристикам близка к реальной (экспериментально снятой с действующего объекта) последовательности. В качестве такой последовательности, например, может быть сетевой трафик.

При проектировании систем получили широкое распространение статистические имитационные модели, воспроизводящие с помощью ЭВМ процесс

функционирования системы или ее части с учетом случайных факторов на базе метода статистических испытаний. Имитационные модели позволяют рассматривать процессы, происходящие в сложной системе, практически на любом уровне детализации. Кроме того, используя алгоритмические возможности моделирующей ЭВМ, в имитационной модели можно реализовать любой алгоритм управления или функционирования системы. Это явилось причиной того, что метод имитационного моделирования стал одним из основных методов исследования сетей связи. В отличие от аналитических методов при имитационном моделировании структура моделирующего алгоритма слабо зависит от совокупности искомых величин и определяется главным образом строением математической модели исследуемого процесса, объекта. Это объясняет наличие большого числа методов построения математических моделей и средств реализации моделирующих алгоритмов, которые различаются средствами формализации процессов, протекающих в реальной системе.

Основным преимуществом имитационных моделей по сравнению с аналитическими является возможность решения более сложных задач. Имитационные модели позволяют легко учитывать наличие дискретных или непрерывных элементов, нелинейные характеристики, случайные воздействия и др. Поэтому этот метод широко применяется на этапе проектирования сложных систем.

Однако нельзя говорить, что имитационная модель «ближе» к реальному объекту, чем аналитическая. Более того, для проверки правильности построения и функционирования имитационной модели ее необходимо поставить в условия, принятые при конструировании аналитической модели. Совпадение результатов аналитического и имитационного моделирования может в некоторой мере служить проверкой правильности функционирования имитационной модели. Только после такой проверки имитационной модели можно снять ряд дополнительных ограничений и упрощений, принятых при конструировании аналитической модели, с целью выявления наиболее существенных сторон функционирования реальной системы не отраженных в аналитической форме.

Физические модели воспроизводят определенные геометрические, физические, динамические либо функциональные свойства проектируемой системы и основаны на использовании принципа аналогии между механическими, электрическими и другими физическими явлениями, описываемыми одинаковыми математическими соотношениями. Если природа элементов физической модели отличается от природы элементов реальной системы, то соответствующая модель называется *аналоговой*. Если физическая модель воспроизводит систему с сохранением ее физической природы и закономерностей, то она называется *натурной*. Примерами натуральных моделей являются макетные и опытные образцы проектируемой системы. Естественно, что натурные модели наиболее точно воспроизводят анализируемую систему (практически являются изоморфными системе), однако требуют значительных затрат на создание, что исключает возможность их применения для анализа множества альтернатив построения системы.

В общем случае на различных стадиях разработки системы осуществляется построение как абстрактных, так и физических моделей. На ранних стадиях разрабатываются аналитические модели для всех рассматриваемых альтернатив построения системы. По результатам анализа с помощью аналитических моделей из рассматриваемого множества альтернатив. Исключаются явно неперспективные альтернативы. Для оставшихся альтернатив осуществляется построение более сложных и точных численных моделей, с помощью которых выбирается, как правило, одна оптимальная по заданному критерию альтернатива. Дальнейший анализ этой альтернативы проводится на натуральных моделях.

Сеть связи является сложной многопараметрической системой, что определяет множественность постановки задачи ее оптимального проектирования (синтезе). Трудность постановки и решения такой задачи обусловили разработку целого ряда математических моделей сети, каждая из которых позволяет решить какую-либо частную задачу по выбранному критерию (задача выбора пропускных способностей каналов сети, задача выбора потоков в каналах и задача выбора топологии сети). При этом используется, как правило, одномерный критерий оптимизации (скалярная оптимизация) и целый ряд допущений.

Таким образом, на сегодня важнейшей задачей остается разработка комплексной математической модели мультисервисной сети и, в частности ее инфотелекоммуникационной транспортной системы, позволяющей связать экономические и структурные характеристики сети с характеристиками процесса доставки сообщений, надежности, управления и технического обслуживания, а также разработка на основе этой модели методов проектирования сетей указанного класса с использованием средств автоматизированного проектирования и системного подхода ко всему циклу ее разработки.

1.6. Проверка и применение моделей

Выбор типа моделей и их построение являются неформальным, творческим процессом, поэтому возникает необходимость проверки достоверности моделей и получаемых с их помощью оценок характеристик и критерия. Рекомендуется проводить следующие виды проверок построенной модели: на *непротиворечивость, чувствительность, реалистичность и работоспособность*.

Проверка на *непротиворечивость* заключается в проверке правильности реакции модели на изменения учитываемых ею параметров. Проверяемая модель считается непротиворечивой, если полученные с ее помощью для граничных значений параметров результаты не противоречат имеющимся опытным фактам, логике или здравому, смыслу.

Проверка на *чувствительность* заключается в оценке относительных изменений характеристики или критерия при изменении параметров в области допустимых значений. Модель является чувствительной по некоторому параметру, если его изменение в области допустимых значений вызывает существенные с точки зрения проектировщика изменения соответствующей характеристики или критерия (рис. 2.2). Если модель оказывается нечувствительной по некоторому параметру, то это означает, что этот параметр является несущественным и может быть исключен из вектора параметров модели.

Проверка на *реалистичность* заключается в проверке соответствия модели известным частным случаям, достоверность которых установлена теоретическим или эмпирическим путем.

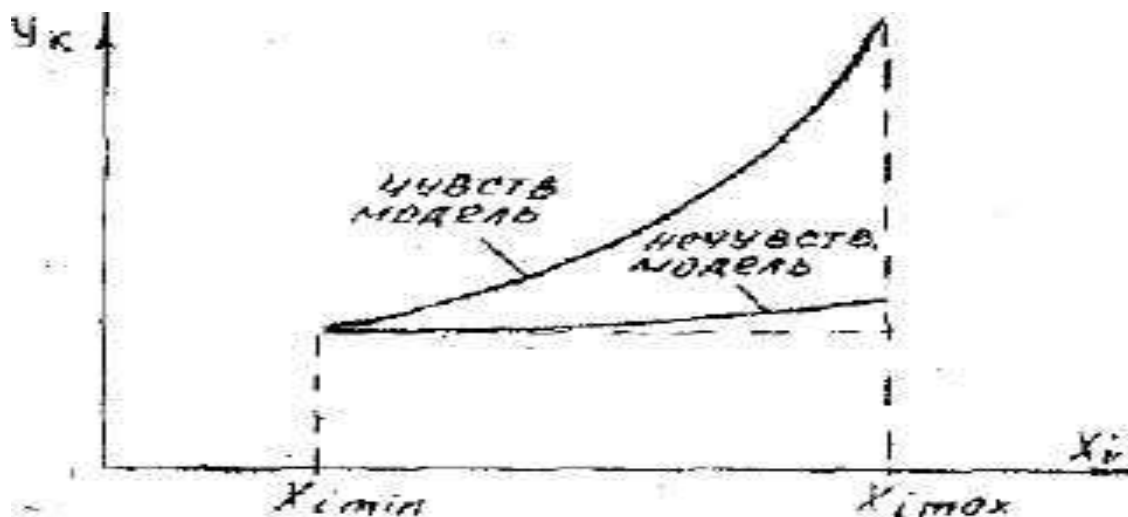


Рис. 1.3 - Проверка модели на чувствительность

Целью проверки модели на *работоспособность* является получение ответа на вопрос о практической применимости предлагаемой модели, успешно прошедший все предыдущие виды проверок, с точки зрения затрат времени, трудовых, материальных и финансовых ресурсов на проведение с ее помощью необходимых расчетов или экспериментов. Положительный ответ на этот вопрос позволяет приступить к применению построенной модели для оценки характеристик и критерия оптимальности системы. Применение модели при анализе системы заключается:

- в планировании выполняемых на ее основе расчетов или экспериментов (выбор вариантов исходных данных и последовательности вариантов расчетов или экспериментов);
- в проведении расчетов или экспериментов на основе модели с помощью ЭВМ в соответствии с подготовленным планом;
- в обработке полученных результатов расчетов или экспериментов (оценка достоверности результатов, построение графиков и т. п.).

При оптимальном синтезе системы применение модели является частью процедуры принятия оптимального решения.

1.7. Функциональная модель сети связи и ее элементов

Функциональные модели сети связи и ее элементов служат для составления на их базе аналитических моделей соответствующих объектов, учитывающих, на выбранном уровне описания, все необходимые преобразования, которые претерпевает сообщение при прохождении через моделируемую систему.

Основное внимание мы уделим описанию процессов, протекающих в системе на уровне «узел коммутации - сеть связи» и составлению функциональных моделей именно этих объектов.

Функциональная модель не преследует цели буквального (алгоритмического) описания процессов, протекающих в системе при проводке по сети и ее элементам каждого отдельного сообщения, однако при ее составлении должны учитываться общая «физическая конструкция» моделируемой системы и порядок взаимодействия ее отдельных элементов при прохождении через систему, потоков сообщений.

Степень подробности описания диктуется физическими соображениями и возможностями применения, различных аналитических методов исследования. Укажем уровни описания и характер решаемых задач, которые необходимо принимать во внимание при построении функциональных моделей сети и ее элементов.

На самом низком уровне, обычно, рассматриваются вопросы моделирования среда распространения сигналов, построение оптимальных методов передачи и приема сигналов в соответствующих средах, выбор формы сигналов, методов регенерации, компенсации искажений, поэлементной синхронизации, изучаются возможности организации приема «в целом» (с введением избыточности непосредственно в передаваемый сигнал), и даются рекомендации по конструкции модемов.

На втором уровне изучаются дискретные каналы (двоичные, m -ичные, кодовые, системные каналы) связи, строятся соответствующие модели дискретных каналов, оцениваются их параметры по реальной статистике, изучаются кодовые методы защиты дискретной информации от ошибок, в частности, методы с применением обратной связи, строятся модели соответствующих систем передачи, рассматриваются и детализируются алгоритмы к каналам и трактам передачи и т.д. На этом уровне (в сочетании с первым уровнем описания) в основном исследуются методы получения максимальной скорости передачи при заданной достоверности.

На третьем уровне рассматриваются вопросы прохождения потоков сообщений через отдельные элементы сети (маршрутизаторы, коммутаторы, концентраторы, абонентские пункты, тракты передачи и т. д.) и всю сеть в целом.

Основная задача исследований на этом уровне - это получение вероятностно - временных характеристик сети и ее элементов (распределение задержек сообщений, эффективная скорость передачи в тракте, вероятность переполнения буферной памяти узлов коммутации (УК), вероятность блокировки и т.д.) в процессе ее функционирования. Основным методом исследования - **теория массового обслуживания**.

На четвертом уровне изучаются общесетевые вопросы, связанные с топологией сети, распределением потоков на сети, управлением сетью и т.д.

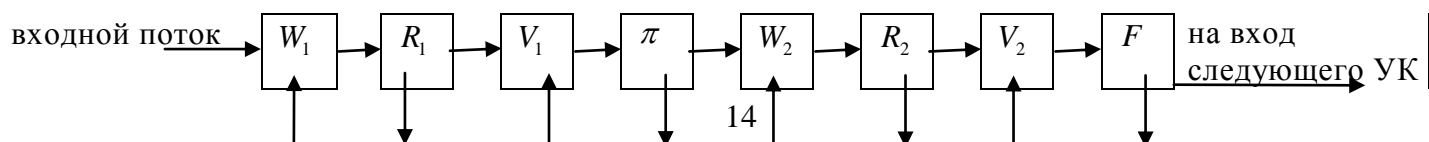
Отдельно стоят вопросы организации программного обеспечения УК, описания его структуры и взаимодействия с аппаратурным комплексом узла.

При построении функциональной модели, прежде всего, производят идентификацию моделируемой системы и соответствующей системы массового обслуживания, т. е. выделяют приборы обслуживания, буферные накопители, фиксируют дисциплины занятия обслуживающих приборов, порядок их освобождения, характер управления очередями, формулируют требования к описанию состояния системы и системы приоритетов.

При моделировании УК как разомкнутой системы массового обслуживания, в качестве приборов обслуживания могут выступать связное оборудование (связной процессор), коммутационное оборудование (соответствующий вычислительный комплекс и программное обеспечение), входящие и исходящие каналы связи, вводно-выводные устройства, отдельные компоненты используемого вычислительного комплекса и т.д.

Характеристики входящих каналов связи (каналов передачи данных) на соответствующем уровне описания их работы могут учитываться или при описании входящих потоков информации (как бы «модулировать» входящие потоки) или вводиться в описание состояния системы при составлении матрицы плотностей перехода.

Если учитывать конечность вычислительной скорости коммутационного процессора (коммутатора), то типичная (и достаточно общая) функциональная модель элемента сети - «узел коммутации + исходящие каналы» (как разомкнутой системы массового обслуживания) будет выглядеть так, как это представлено на рис.1.4.



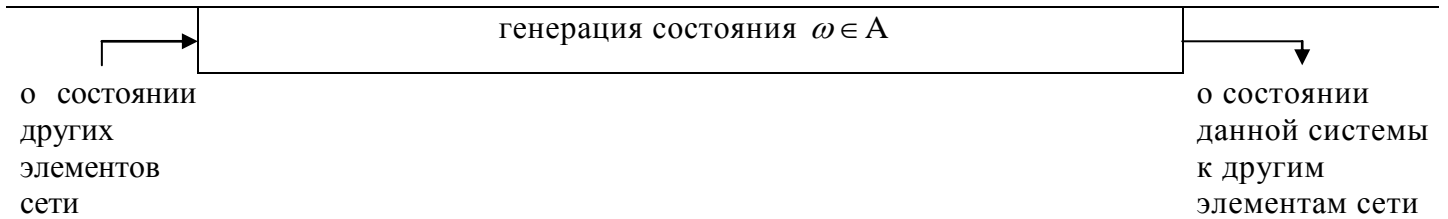


Рис. 1.4 - Функциональная модель элемента сети «узел коммутации + исходящие каналы»

Эта же схема при соответствующих упрощениях может служить функциональной моделью концентратора, ретранслятора, оконечной системы (*End System, ES*) и т. д.

Здесь обозначено $W_1 \equiv W_1(\omega) = [w_{ij}^{(1)}]$ - стохастическая матрица размера $l \times l$, где $w_{ij}^{(1)}$ - вероятность того, что в состоянии $\omega \in A$ элемент коммутации (заявка) из i -го входящего канала (или заявка i -го приоритета) будет направлена в j -ю очередь первого накопителя;

l - число входящих каналов (или приоритетов);

R - объем накопителя на входе УК;

$V_1 \equiv V_1(\omega) = [v_{ij}^{(1)}]$ - стохастическая матрица размера $l \times l$, где $v_{ij}^{(1)}$ - вероятность того, что в состоянии ω на j -й вход 1-го обслуживающего прибора УК при его освобождении поступит заявка из i -й очереди;

Сетевой план распределения информационных потоков в сети с пакетной коммутацией задается применительно к каждому УК (маршрутизатору, коммутатору) при разработке общего плана распределения потоков в сети исходя из средних характеристик сети и ее пользователей. Эта информация, обычно, оформляется в каждом узле в виде матрицы коммутации. Матрица коммутации есть дешифратор, связывающий адрес получателя с номером исходящего канала. Аналитически статическая сетевая коммутация (план распределения информации) задается в виде стохастической матрицы $\pi = [\pi_{ij}]$, размера $l \times m$, где l - число входящих каналов, m - число исходящих каналов, являющейся произведением двух матриц S_1 и S_2 соответственно размера $l \times 2^n$ и $2^n \times m$, где n - число разрядов адресной части сообщения (адрес получателя). Элемент (ij) матрицы S_1 есть вероятность появления в i -том входящем канале (очереди) сообщения с адресом j ($j \in 0, 2^n - 1$) через j -ый выход. Матрица S_2 - есть матрица коммутации, (ij) -й элемент которой есть вероятность выдачи обслуженной (скоммутированной) заявки (сигнального сообщения, пакета и т. д.) с адресом i ($i \in 0, 2^n - 1$) через j -ый выход. При пакетной коммутации (ij) -м элементом матрицы коммутации может быть вероятность посылки сообщения с адресом i ($i \in 0, 2^n - 1$) в j -ый исходящий канал. $\pi = S_1 H S_2$, $H = \text{diag} \{H_0(t), H_1(t), \dots, H_{2^n-1}(t)\}$, $H_k(t)$ - вероятность того, что элемент коммутации с

адресом получателя k ($k \in \overline{0, 2^n - 1}$) будет обрабатываться в УК за время равное или меньшее t .

В частности, в случае экспоненциального времени обработки адреса k с интенсивностью обслуживания ν_k величина $\pi_{ij} \Delta t = \Delta t \sum_{k=0}^{2^n-1} s_{ik}^{(1)} \nu_k s_{kj}^{(2)}$ - есть вероятность того, что за время Δt 1-й обслуживающий прибор (коммутатор) освободится через j -й выход от заявки, поступившей на i -й вход (здесь $s_{ik}^{(1)}$ и $s_{kj}^{(2)}$ - элементы матриц S_1 и S_2). Если пренебречь временем обработки адреса в УК, то следует положить $\Delta t \nu_k \equiv 1, (k \in \overline{0, 2^n - 1})$.

Для различных УК матрицы $\pi = [\pi_{ij}]$ различны. На УК с технологией коммутации каналов матрица π отсутствует;

$W_2 \equiv W_2(\omega) = [w_{ij}^{(2)}]$ стохастическая матрица размера $m \times m$, где $w_{ij}^{(2)}$ - вероятность того, что в состоянии ω заявка с i -го выхода 1-го обслуживающего прибора поступит в j -ю очередь 2-го накопителя;

$V_2 \equiv V_2(\omega) = [v_{ij}^{(2)}]$ - стохастическая матрица размера $m \times m$, где $v_{ij}^{(2)}$ - вероятность того, что в состоянии ω заявка с i -й очереди второго (выходного) накопителя поступит в j -ый исходящий канал;

$F = [F_1(t), F_2(t), \dots, F_m(t)]$, где $F_i(t)$ - вероятность того, что элемент коммутации (заявка) будет обрабатываться (передаваться) в i -м исходящем канале время равное или меньше t ; В частности, в случае экспоненциального времени обработки (передачи) под F будем подразумевать вектор $[\mu_1, \mu_2, \dots, \mu_m]$, где μ_i - интенсивность обслуживания (передачи) элемента коммутации в i -м исходящем канале.

Векторный входящий поток может быть дискретным или непрерывным, ординарным или неординарным.

Наиболее типичным ординарным и непрерывным во времени процессом является пуассоновский с параметрами $\lambda_1, \lambda_2, \dots, \lambda_l$, где λ_i - интенсивность появления элементов коммутации на входе УК в i -м канале; Наиболее типичным неординарным и дискретным - является биномиальный канал с параметрами p_1, p_2, \dots, p_l , где p_i - вероятность появления на входе УК элемента коммутации в i -м канале (или вероятность появления на входе УК элемента коммутации с i -м приоритетом).

Отметим, что исследование моделей систем передачи, в состав, которых входят ненадежные приборы обслуживания (приборы с очередями) может быть сведено к исследованию моделей систем с приоритетами, т.к. отказ прибора может рассматриваться как заявка, обладающая наивысшим абсолютным приоритетом.

Если УК рассматривается как элемент сети, то поток на его входе состоит из двух компонент: *потока транзитных элементов коммутации*, который может существенно отличаться от пуассоновского или биномиального, и *потока, порождаемого конечными системами ES*, подсоединенными к данному УК. Аналогичную структуру имеет и выходной поток УК.

Необходимо отметить, что исследования различных типов сетевого трафика за последние полтора десятка лет доказывают, что сетевой трафик является самоподобным (self-similar) или фрактальным (fractal) по своей природе. «Самоподобие» представляет собой свойство процесса сохранять свое поведение и внешние признаки при рассмотрении в разном масштабе. Из этого следует, что

используемые методы моделирования и расчета сетевых систем, основанные на использовании пуассоновских потоков, не дают полной и точной картины происходящего в сети. На сегодняшний день разработано множество моделей, предназначенных для имитации фрактального трафика: *фрактальное броуновское движение (Fractional Brown Motion - FBM)*; *фрактальный гауссовский шум (Fractional Gaussian Noise - FGN)*; *хаотические отображения (Chaotic Map – CMAP)*; *модели на основе техники "динамического моделирования Маркова" (Dynamic Markov Modelling – DMM)*; *модели с использованием нечеткой логики*; *нейросетевые модели*; *авторегрессионные модели (Autoregressive Models - AR)*; *Фрактальные точечные процессы (Fractal Point Process – FPP)*; *ON/OFF – модели*; *Фрактальное движение Леви движение (Fractional Levi Motion - FLM)*; *Мультифрактальные модели (Multifractal - MF)*; *вейвлет модели (Wavelet Models)*; *Модели на основе классических систем массового обслуживания*. Как правило, такие модели удачно описывают трафик с пуассоновскими потоками. Однако, такая модель как *M/G/N* способна создать приблизительно самоподобный трафик путем управления поведением «хвоста» произвольного распределения обслуживания пользователей, создавая тем самым долговременную зависимость.

Для моделирования группового потока с эффектом самоподобия хорошо подходит модель, базирующаяся на применении фрактального Броуновского движения [9 Агеев Д.В., Игнатенко А.А., Копылев А.Н. Методика определения параметров потоков на разных участках мультисервисной телекоммуникационной сети с учетом эффекта самоподобия [Электронный ресурс] // Проблемы телекоммуникаций. – 2011. – № 3 (5). – С. 18 – 37. – Режим доступа: http://pt.journal.kh.ua/2011/3/1/113_ageyev_method.pdf.]. Эта модель предложена Leland и др. [10 Leland W.E., Taqqu M.S., Willinger W., Wilson D.V. On the Self-Similar Nature of Ethernet Traffic (Extended Version) // IEEE/ACM Trans, on Networking. – 1994. – Vol. 2, Issue 1. – P. 1–15.] и использована Норросом [6, 7 Norros I. A Storage Model with Self-Similar Input // Queueing Systems. – 1994. – Vol. 16, No 3-4. – P. 387–396. Norros I. On the use of fractional Brownian motion in the theory of connectionless networks // Selected Areas in Communications, IEEE Journal. – 1995. – Vol. 13, Issue 6. – P. 953–962.] для получения выражений, описывающих качественные характеристики обслуживания группового самоподобного потока в узлах сети.

Длина передаваемых пакетов (сообщений) может либо учитываться непосредственно при построении моделей входящих потоков (например, моделирование входящих потоков альтернирующими процессами восстановления, полумарковскими процессами или пуассоновскими точечными процессами с сопровождающими их потоками длин пакетов (сообщений) либо трактоваться как часть времени обслуживания элементов коммутации в обслуживающих приборах (УК, исходящие каналы).

При *коммутации каналов* блоки R_2 (память на выходе) и W_2 (управление очередями на выходе) отсутствуют. Блоки V_1, π, V_2 сливаются в единое устройство управления (которое может рассматриваться как прибор обслуживания или «фаза» обслуживания) и мы приходим к обычной модели однофазной многолинейной системы массового обслуживания.

В системах с *коммутацией пакетов* приборами обслуживания являются как УК, так и исходящие каналы. В этом случае элемент сети – узел коммутации является двухфазной многолинейной системой обслуживания. Свободные исходящие каналы в УК здесь могут предоставляться или в заранее определенном порядке (таблице коммутации в режиме виртуального канала) или в случайном порядке – в датаграммном режиме. Дисциплина выбора ожидающего элемента коммутации (заявки) из некоторой очереди в накопителе и дисциплина выбора самой очереди при наличии нескольких очередей также может быть

различной. Если дисциплина выбора из очереди, как правило, не оказывает влияния на финальный вектор вероятностей системы, то дисциплина выбора между очередями влияет на финальные вероятности. Так, очевидно, что вероятность переполнения накопителя (вероятность потерь в системе с КК) будет меньше, если дисциплина выбора между очередями при прочих равных условиях определяется их длиной. Введение приоритетов для входящих потоков, динамическое распределение памяти, может также заметно повлиять на вероятность переполнения накопителей или среднее время задержки сообщений.

Если предположить, что процессы коммутации протекают мгновенно (по сравнению со скоростью передачи информации в системе), то данная модель упрощается до однофазной (исчезает или память на входе или память на выходе).

Функциональная модель всей сети, которая является в принципе многофазной и многолинейной частично замкнутой системой массового обслуживания, образуется из рассмотренных выше ячеек согласно матрице смежности сети (рис.1.5).

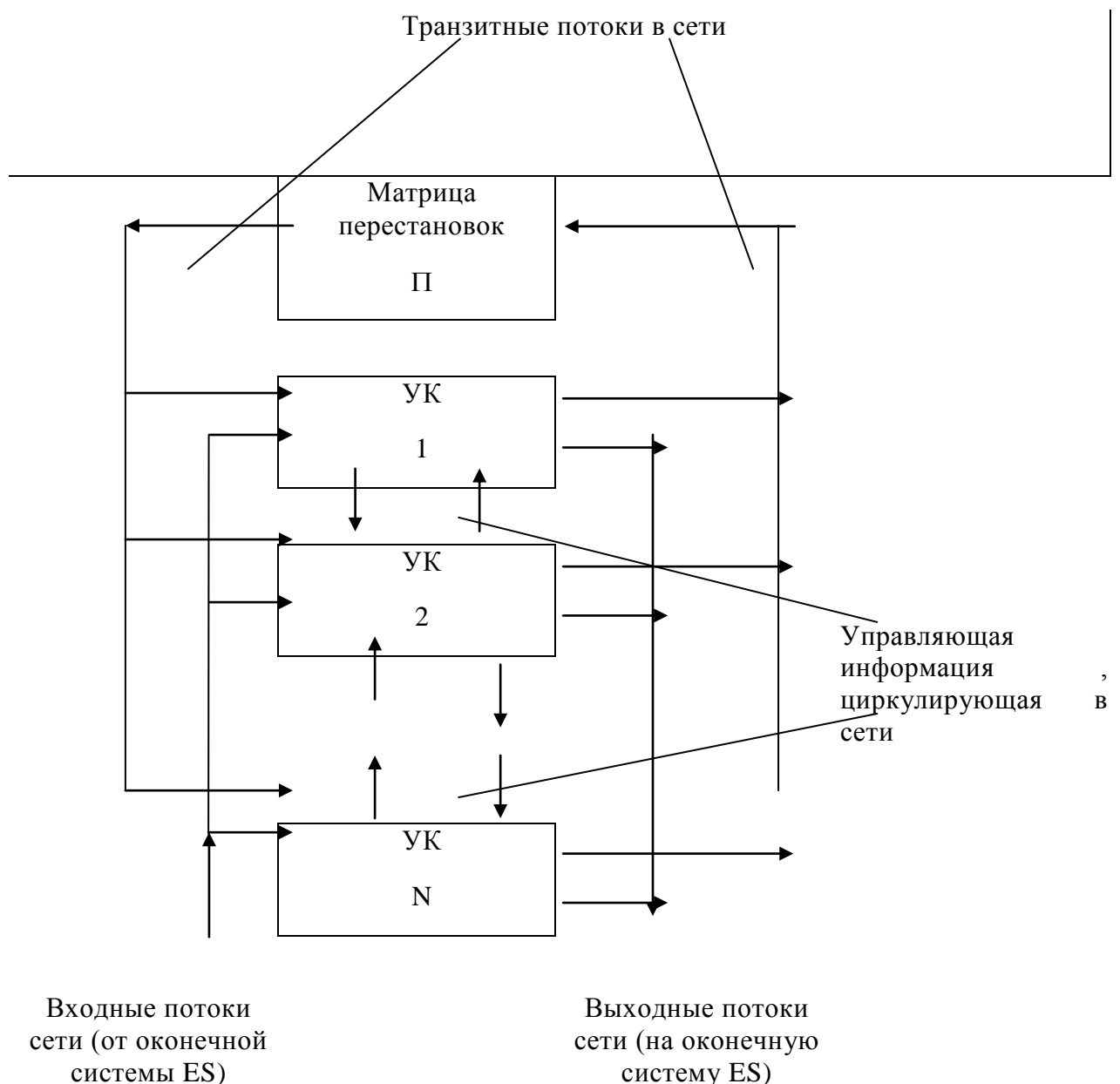


Рис. 1.5 – Функциональная модель сети

Матрица перестановок Π , размер которой задается числом транзитных каналов сети, однозначно определяется матрицей смежности сети (матрица размера $N \times N$ с единицей на пересечении i -й строки и j -го столбца, если i -й и j -й УК соединены каналами, и нулем в противном случае; N - число УК в сети) и указывает на номера выходных и входных (относительно УК) транзитных каналов сети, соединяемые согласно топологии сети.

Как следует из функциональной модели сети, представленной на рис.1.5, обслуженные заявки (элементы коммутации) или возвращаются с некоторыми вероятностями на входы других коммутационных центров или выводятся из системы.

Потоки на входах УК состоят из потоков заявок, вводимых в сеть, и заявок прошедших транзитные узлы (т.е. возвращенных на входы соответствующих УК).

При моделировании УК с пакетной коммутацией более точный учет структуры программного обеспечения вычислительных средств узла приводит к распределению времени обслуживания, образованному двумя величинами: постоянной составляющей и случайной составляющей, имеющей при оптимальном выборе структуры и системы команд коммутационных процессоров распределение, близкое к экспоненциальному.

Второй прибор обслуживания (исходящие каналы) также имеет распределение, являющееся сверткой двух величин постоянной (задержка, связанная со временем распространения сигналов) и случайной, обусловленной возможными переспросами в каналах и трактах передачи данных. Это распределение образуется на уровне элемента защиты информации от ошибок (блок, пакет, если элемент защиты совпадает с элементом коммутации) и зависит от характеристик каналов передачи данных и алгоритмов обмена в тракте передачи (алгоритмы с ожиданием, алгоритмы с решающей обратной связью и блокировкой, адресные алгоритмы и т.д.). В первом приближении можно считать, что время обслуживания на второй фазе имеет либо вырожденное распределение (время обслуживания постоянное) либо распределение, близкое к экспоненциальному (в случае непрерывной во времени модели) или геометрическому (в случае дискретной во времени модели) с параметрами, определяемыми средними временами задержки сообщения в каналах.

Вопросы для самопроверки по теме 1.

1. Сформулируйте принцип системного подхода к проектированию сложных систем.
2. Сформулируйте основные процедуры проектирования систем. В чем состоят назначение и отличие процедур синтеза и анализе системы?
3. Что понимается под параметрами, характеристиками и критерием оптимальности системы?
4. Сформулируйте в общем виде задачу оптимального проектирования системы.
5. Сформулируйте задачу параметрического синтеза для альтернативы $A_i \in A$ при оптимизации по максимуму критерия эффективности.
6. Какие методы могут применяться для решения задачи оптимального параметрического синтеза, их особенности?
7. Сформулируйте задачу структурного синтеза системы.
8. Какие виды моделей применяются в процессе проектирования ИКС их достоинства и недостатки?
9. В чем заключается проверка и применение моделей?
10. Опишите функциональную модель элемента сети «узел коммутации + исходящие каналы»
11. Как строится функциональная модель всей сети?

ТЕМА 2. ИНФОТЕЛЕКОММУНИКАЦИОННАЯ ТРАНСПОРТНАЯ СИСТЕМА КАК ПРЕДМЕТ МОДЕЛИРОВАНИЯ

3.1. *Инфокоммуникационная сеть и ее компоненты*

Услуги глобальной информационной инфраструктуры будут предоставляться посредством дистанционного доступа, организуемого на основе инфокоммуникационных сетей (ИКС) с архитектурой сетей следующего поколения NGN (Next Generation Networks), подчеркивая конвергенцию информационных и телекоммуникационных компонентов в формировании новых сетевых услуг на основе мультисервисной платформы.

Идея разработки NGN была предложена в 2001 г. Европейским институтом стандартов электросвязи ETSI (European Telecommunications Standards Institute) и поддержана Сектором стандартизации телекоммуникаций Международного союза электросвязи (МСЭ-Т). Первые две рекомендации МСЭ-Т - Y.2001 и Y.2011 - были утверждены в конце 2004 г. в новой серии Y. 2000, специально выделенной для рекомендаций о NGN. На начало 2011 года в этой серии уже было 70 рекомендаций, которые относятся к так называемой первой версии NGN (NGN release 1). В последнее время в МСЭ-Т начаты работы по второй версии (NGN release 2).

Согласно определению, приведенному в Рекомендации МСЭ-Т Y.2001, сеть следующего поколения NGN – это сеть с *пакетной коммутацией*, способная обеспечить пользователей разнообразными узкополосными и широкополосными услугами, включая услуги телефонной связи, основанная на широкополосной сети с пакетной технологией транспортировки, обеспечивающей необходимое качество услуг QoS (Quality of Service), в которой функции, связанные с предоставлением услуг, не зависят от технологий транспортировки информации. Сеть NGN дает пользователям неограниченный доступ к различным услугам провайдеров и поддерживает обобщенную мобильность, которая позволяет пользователям получить доступ к услугам в любом месте и в любое время.

В рекомендации МСЭ-Т Y.2012 перечислены основные принципы функциональной архитектуры NGN:

1. Поддержка множества технологий доступа – функциональная архитектура NGN должна обладать гибкой конфигурацией, необходимой для поддержки множества технологий доступа.
2. Распределенное управление – должен использоваться принцип распределенной обработки в пакетных сетях и поддерживаться прозрачность местоположения для распределенных вычислений.
3. Открытое управление – сетевые интерфейсы управления должны быть открыты для поддержки процессов создания новых и изменения существующих услуг и поддержки средств обеспечения логики услуг сторонних поставщиков.
4. Независимость предоставления услуг – процесс предоставления услуг должен быть разделен между функциями транспортной сети, работающей с использованием указанного выше механизма распределенного открытого управления. Это приведет к поддержке конкурентного окружения при развитии NGN, которое будет способствовать ускорению процессов внедрения новых услуг.
5. Поддержка услуг конвергентных сетей - это необходимо для создания гибких, простых в использовании мультимедийных услуг для замещения технических возможностей конвергентных фиксировано-мобильных сетей с помощью функциональной архитектуры NGN.
6. Расширенные возможности безопасности и защиты – это базовый принцип открытой архитектуры, он требует обязательной защиты сетевой инфраструктуры с помощью механизмов обеспечения соответствующих уровней безопасности и живучести сети.

Для реализации этих функций в Рекомендации МСЭ-Т Y.2011 предложена базовая эталонная модель NGN, включающая два уровня: уровень услуг NGN (service stratum) и уровень транспорта NGN (transport stratum), каждый из которых содержит по три плоскости: пользователя, управления и менеджмента (рис. 3.1).

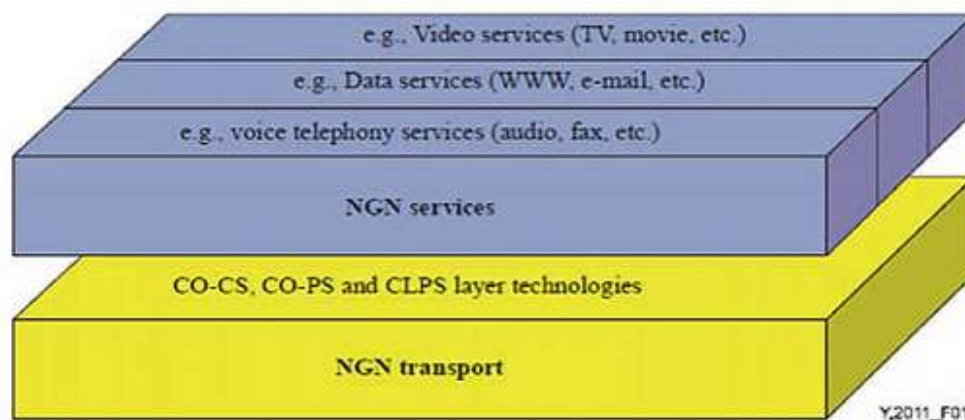


Рис. 3.1. Базовая эталонная модель NGN

В рекомендациях Международного союза электросвязи (МСЭ/ITU) дано следующее определение Next Generation Network: «NGN — это сеть с коммутацией пакетов, способная предоставлять телекоммуникационные услуги посредством широкополосных транспортных технологий, поддерживающих качество обслуживания (QoS), в которой функции услуг не зависят от используемых транспортных технологий».

Отличительной чертой модели NGN, предлагаемой сектором МСЭ-Т, является её функциональное деление на уровень услуг и транспортный уровень. Последний обеспечивает выполнение функции обмена дискретной информацией любого типа между любыми двумя географически разнесёнными точками.

Первый уровень реализует прикладные функции, связанные с востребованными услугами, например с организацией передачи речи и видеоизображений по отдельности или в комбинации. В соответствии с рекомендациями МСЭ-Т, NGN должна осуществлять конвергенцию услуг передачи данных, речи, видео-, аудио- и визуальных данных в индивидуальном, групповом и широковещательном режимах.

С целью более простого понимания принципов построения сетей следующего поколения приведем обобщенную 4-х уровневую архитектуру NGN, в которой выделяются следующие уровни (рис. 3.2). - слой доступа, содержащий сеть абонентского доступа к транспортной пакетной сети;

- транспортный слой, включающий магистральную пакетную сеть, отвечает за гарантированную передачу информации через транспортную сеть с различным уровнем качества. Они обеспечивают механизмы реализации заданного уровня качества передачи QoS для пользовательского трафика включая управление буферами, очередями и расписанием, фильтрацию пакетов, классификацию, маркирование и формирование трафика, контроль соблюдения правил обслуживания, управление шлюзами и функции межсетевых экранов.

- слой управления коммутацией, включает совокупность функций по управлению всеми процессами обслуживания вызовами в телекоммуникационной сети;

- слой услуг и эксплуатационного управления, который содержит логику выполнения услуг и/или приложений и управляет этими услугами, имеет открытые интерфейсы для использования сторонними организациями (для разработки программ и новых услуг).

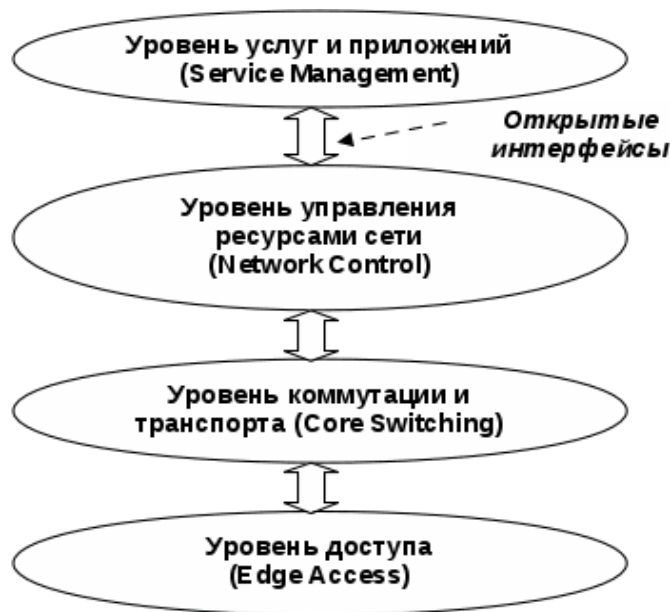


Рис.3.2 - Обобщенная 4-хслойная архитектура NGN

Терминальное оборудование не входит в состав сети NGN и в принципе может быть любым из набора абонентского оборудования существующих проводных и беспроводных сетей. Однако такое терминальное оборудование может быть включено в сеть NGN только через согласующее иллюзовое абонентское оборудование уровня доступа. Непосредственное подключение к сети возможно только пакетных абонентских терминалов, работающих с использованием протоколов SIP и H.323.

Каждый слой охватывает один или несколько уровней модели OSI, причем каждый уровень концептуально состоит из плоскости данных (или плоскости пользователя), плоскости управления и плоскости менеджмента.

3.2. Мультисервисная сеть связи как область взаимодействия инфокоммуникационной сети NGN

Область взаимодействия инфокоммуникационной сети NGN в терминах модели взаимодействия открытых систем образует ее цифровая телекоммуникационная **мультисервисная сеть связи (МСС)**, в состав которой входят абоненты мультисервисной сети, широкополосные сети доступа и транспортный слой или уровень.

Структура МСС показана на рис.3.3. МСС – это сеть связи, построенная в соответствии с концепцией NGN и обеспечивающая предоставление **мультимедиа**. Мультимедиа означает интеграцию нескольких информационных типов сообщений, таких как текст, изображения, графика, анимация и многое другое. Создание мультимедиа стало главным направлением в развитии информационных технологий последнего десятилетия и привело не только к появлению новых технологий, но и возникновению новых сервисов.

Основные сетевые характеристики интегрального обслуживания трафика различной природы в МСС обеспечивает ее пакетное коммуникационное мультипротокольное ядро или пакетная мультипротокольная **инфотелекоммуникационная транспортная система (ИТС)** [14], которая реализует услуги переноса (bearer service) мультимедийной информации между сетевыми окончаниями (Network Terminator, NT) без какого-либо анализа или обработки ее содержания с учетом заданного качества обслуживания QoS.

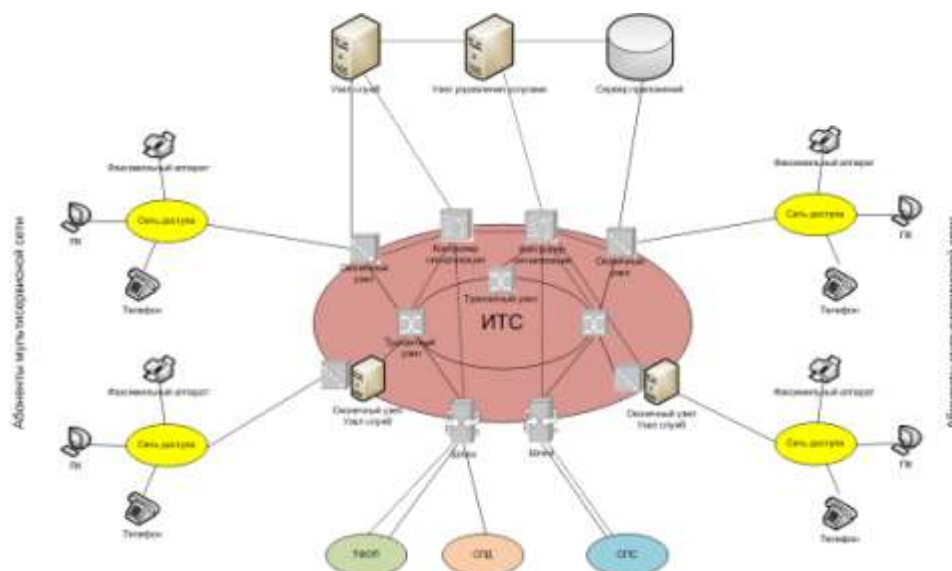


Рис. 3.3 – Структура инфокоммуникационной сети NGN

ИТС обеспечивает перенос разных видов информации с использованием различных протоколов передачи. Услуги переноса (доставки) информации характеризуются:

- типами соединений (Connection Type, CT);
- классом качества услуги (Class of Service, CoS);
- параметрами трафика (Traffic Parameters, TP).

ИТС по сути является транспортным слоем ИТС NGN.

Доступ к ресурсам мультисервисной сети осуществляется через граничные узлы, к которым подключается оборудование сети доступа или осуществляется связь с существующими сетями. В последнем случае граничный узел выполняет функции межсетевого шлюза. Под сетью доступа подразумевается системно-сетевая структура, состоящая из абонентских линий, узлов доступа и систем передачи. Она служит для организации подключения пользователей к ресурсам региональных сетей. Услуги, предоставляемые сетью доступа, можно различать по способам доставки информации, качеству услуги (QoS) и скорости передачи.

В рамках концепции Взаимовязанной сети связи (ВСС) Российской Федерации [2] национальна мультисервисная сеть будет базироваться на двухуровневой архитектуре (рис. 3.4). На региональном уровне мультисервисная сеть должна обеспечивать подключение абонентов и предоставление им как транспортных, так и инфокоммуникационных и других услуг, а также обеспечивать возможность взаимодействия с аналогичными услугами других региональных сетей.

На магистральном уровне мультисервисная сеть должна обеспечивать предоставление услуг переноса для взаимодействия мультисервисных региональных сетей, а также для передачи (при необходимости) нагрузки всех существующих сетей. Конечными целями построения национальной ВСС являются [2]:

- обеспечение широкого спектра инфоуслуг с возможностью доступа к глобальным информационным ресурсам;
- организация широкополосного абонентского доступа;
- предоставление мультимедийного транспортного сервиса единым образом в общей широкополосной физической среде с требуемым качеством обслуживания QoS;
- повышение эффективности использования сетевых ресурсов на коллективной основе;

- обеспечение широкого диапазона гибкости и многофункциональности терминальных мультимедийных соединений;
- объединение и автоматизация эксплуатационно-технического обслуживания и административного управления системой в целом;
- обеспечение услуг безопасности в процессе хранения, обработки и передачи информации.

Таким образом, потребности информационного общества ставят на повестку дня необходимость решения следующих научных и технических проблем:

- создание региональных и магистральных компонент МСС;
- обеспечение широкополосного цифрового абонентского доступа к инфоуслугам сети на базе набора стандартных широкополосных интерфейсов;
- создание универсальных мультисервисных терминалов с возможностью автоматического выбора конкретного пользовательского сервиса из предлагаемого списка;
- обеспечение защиты мультимедийной информации при ее передаче, обработке и хранении.

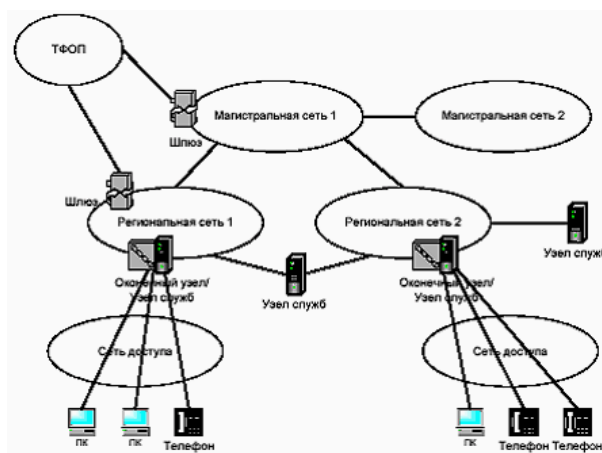


Рис. 3.4. Двухуровневая архитектура мультисервисной сети связи

В нашей стране создание МСС находится на стадии разработки концепций, планирования и проработки вариантов построения [2]. Поэтому важно наличие моделей, методов и алгоритмов анализа сетей указанного класса, на базе которых возможно строить инженерные методики их проектирования. При этом желательно, чтобы эти методики учитывали специфику функционирования ИТС, режимы коммутации и переноса мультимедийного трафика, многообразие сетевых параметров, основные протоколы транспортной архитектуры, а также позволяли оценивать влияния механизмов защиты на их характеристики и ресурсы.

Анализ пакетных технологий, рекомендованных для построения национальных ИКС нового поколения, позволяет сформулировать базовые принципы построения сетей указанного класса.

Принцип QoS-маршрутизации, предполагающий наличие механизмов установления мультимедийного соединения с заданными количественными и качественными параметрами. QoS-маршрутизация обеспечивает: а) формирование и трансляцию сигнального запроса на установление мультимедийных соединений; б) управление процессом резервирования требуемых сетевых ресурсов для мультимедийных соединений на транзитных узлах устанавливаемого маршрута или граничных узлах домена; в) формирование мультимедийных соединений требуемой потоковой структуры в рамках единой транспортной услуги с фиксацией маршрута передачи информации любой

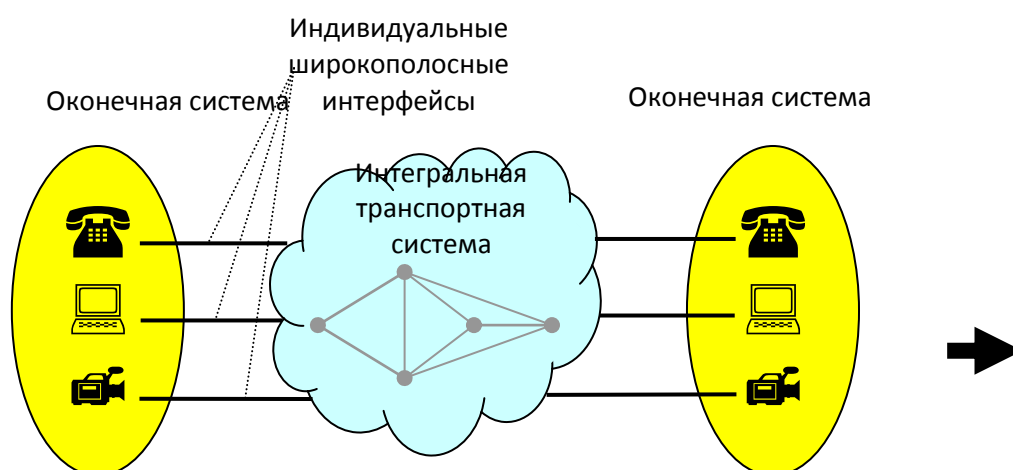
природы.

Принцип масштабируемости сетевых ресурсов, предписывающий наличие механизмов контроля допустимости резервирования требуемых сетевых ресурсов на фазе установления мультимедийной сессии, которые обеспечивают резервирование необходимых сетевых ресурсов или предоставление приложению заранее определенного уровня сервиса CoS, а также формирование маршрутных или коммутационных таблиц.

Принцип «совмещения» разнородного трафика, предписывающий наличие механизмов параметризации («окраски») многокомпонентных информационных потоков с целью их переноса в сессии единым образом в общей физической среде с требуемым качеством обслуживания в рамках единой транспортной услуги.

3.3. *Классификация инфотелекоммуникационных транспортных систем МСС по способу закрепления физических ресурсов сети за соединением*

Вторичные сети связи, классифицируемые по способу коммутации в своем большом разнообразии, фактически характеризуются реализуемой в них телекоммуникационной транспортной системой, которые определяют основные характеристики сетевого обслуживания и рассматриваются в аспекте канального, сетевого и транспортного уровней ЭМ ВОС. Таким образом, для исследования процессов функционирования МСС достаточно ограничиться рассмотрением функций и свойств их ИТС, которые определяют основные характеристики сетевого обслуживания [14]. Выбор конкретных реализаций методов передачи и коммутации (или их комбинаций) диктуется заданными условиями проектирования конкретного типа ИТС и связан с оценкой их эффективности. При этом требования к сетевым алгоритмам, диктуемые характером изохронной нагрузки, всегда являются доминирующими. Анализ специфики функционирования ИТС показывает, что основным назначением ИТС является организация и качественное обслуживание мультимедийного соединения с предоставлением номинальных сетевых ресурсов. Особенностью ИТС является требование обеспечить заданное качество обслуживания мультимедийного соединения (как для изохронного, так и асинхронного трафика), организуемого с мультимедийного терминала в рамках единой транспортной услуги. В этом аспекте ИТС следует рассматривать как новую ступень эволюционного развития интегральных транспортных систем (рис. 3.5).



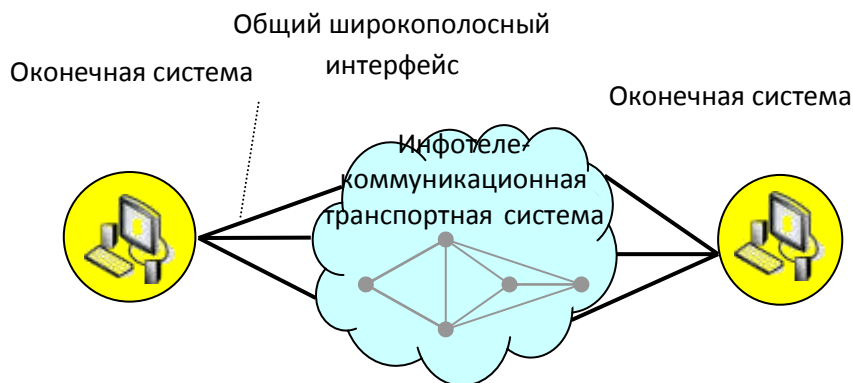


Рис. 3.5. Эволюция интегральных транспортных систем

В целях систематизации моделей, методов и алгоритмов анализа процессов функционирования ИТС и выбора конкретных реализаций с учетом заданных условий проектирования будем классифицировать их на три класса по способу закрепления сетевых ресурсов за соединением в сессии. В этой связи будем различать три типа ИТС: канальную ИТС-К с бесконфликтным или статическим («жестким») закреплением физических ресурсов за соединением, пакетную ИТС-П с динамическим или «нежестким» закреплением физических ресурсов за соединением в режиме коллективного доступа, а также комбинированную (гибридную) ИТС-Г, в которой возможна смешанная стратегия распределения ресурсов. По типу используемых транспортных систем можно классифицировать также и сети связи. Такая классификация является более общей и позволяет объединить большое разнообразие сетей, построенных на одном из базовых методов коммутации и всевозможных его модификациях, в один класс, а также выявить специфику организации переноса информационных потоков и механизмы обеспечения качества их обслуживания в сессии.

Например, в ИКС на технологии АТМ реализовано несколько стратегий закрепления сетевых ресурсов за виртуальным соединением в сессии: при использовании службы СВР (наряду со службами VBR, ABR и UBR) реализуется стратегия с «гибридным» закреплением сетевых ресурсов за соединениями, так как в режиме установленного соединения часть пропускной способности ЛЦТ, которая резервируется за СВР-каналом, не подлежит перераспределению, т.е. полоса пропускания выделяется постоянно не зависит от степени ее реального использования, а перенос изохронного трафика класса А в сеансе связи производится в режиме эмуляции коммутации каналов или с «жестким» закреплением физических ресурсов сети за виртуальным соединением. В то же время при упразднении службы СВР в сети АТМ последняя переходит в категорию чисто пакетных ТС.

3.4. Характеристики базовых потоков информации в ИКС

3.4.1. Переменная и постоянная составляющие сетевой задержки. Способы обеспечения изохронности передачи трафика.

В соответствии с рекомендациями серии I.320 информация, циркулирующая в ИКС, делится на три группы: информация пользователя (группа **U**), информация сигнализации (группа **S**) и информация административного управления (группа **M**). Эти три информационные группы являются базой, с помощью которой и осуществляется доступ к широкополосным пользовательским службам или инфоуслугам ИКС. Предоставление разнообразных инфоуслуг в общем случае сводится к передаче, обработке и хранению элементов (сообщений, пакетов, пакетов) многокомпонентных информационных потоков с заданными количественными (максимальная скорость передачи (Peak rate, PR); средняя скорость передачи (Sustainable rate, SCR); минимальная скорость передачи (Minimum rate, MCR), максимальный размер пульсаций нагрузки

(Maximum burst size, MBS) и др.) и качественными параметрами (постоянная составляющая сетевой задержки (transfer delay, TD) и ее дисперсия или переменная составляющая сетевой задержки (delay variation, DV), доля потерянных элементов потока (loss ratio, LR), достоверность передачи и др.) в рамках мультимедийного соединения [16]. Основные характеристики указанных классов трафика приведены в таблице 3.1

Таблица 3.1- Основные характеристики классов трафика в терминах ATM Forum

Класс трафика	Характеристика
A	Постоянная битовая скорость. С установлением соединения. Требуется изохронность передачи трафика (аудио-, видеоинформация)
B	Переменная битовая скорость. С установлением соединения. Требуется изохронность передачи трафика (уплотненная аудио-, видеоинформация)
C	Переменная битовая скорость. С установлением соединения. Требования к изохронности передачи трафика не предъявляются (неравномерный компьютерный трафик сетей TCP/IP, X.25, frame relay). Требования к достоверности передачи высокие
D	Переменная битовая скорость. Без установления соединения. Требования к качеству передачи трафика не предъявляются (компьютерный трафик UDP, Ethernet, SNMP)

Для унификации и классификации разнородных информационных потоков группы U в дальнейшем изложении будем придерживаться терминов ATM Forum, которым было определено четыре основных классов трафика A, B, C и D на основе следующих признаков: наличие требования организации постоянной или переменной скорости передачи в сеансе связи; наличие требования к изохронности передачи информации; типом протокола организации передачи информации с установлением соединения или без установления соединения.

Трафик (классов A и B) характеризуется низким коэффициентом пульсаций, высокой чувствительностью к задержкам передачи, отражающихся на качестве воспроизводимого непрерывного сигнала, и низкой чувствительностью к потерям информационных элементов, требует режима переноса в сессии, при котором необходимо сохранять с заданной точностью временное расположение элементов потока относительно друг друга. **Свойство сохранять с заданной точностью временное расположение элементов потока относительно друг друга в сессии принято называть изохронностью потока.** Укажем два основных требования, предъявляемые трафиком классов A и B к своей передаче в сеансе связи:

1) поддержание заданной величины **постоянной составляющей сетевой задержки** (network delay, transit delay или latency) элементов потока, определяющей реальное время их доставки;

2) обеспечение заданной величины **переменной составляющей сетевой задержки** или ее флуктуации (jitter) [17], которая определяет требуемый уровень изохронности потока.

Для поддержания непрерывности передачи в реальном времени, цифрового речевого сигнала значение постоянной составляющей сетевой задержки ячеек от абонента до абонента не должно превышать величины порядка 0,3 – 0,5с [17]. Речь традиционно трактуется как трафик от непрерывного источника, имеющий чередующиеся периоды активности и молчания. В этой связи для повышения использования пропускной способности канала связи при передаче речевой информации необходимо учитывать статистику речевых сигналов. Качество восприятия речи не критично к паузам между словами (группами слов) до 300 мс, а для 10% случаев до 1 с [18]. Время задержки не

обязательно должно быть симметричным относительно участвующих в переговорах абонентов. Однако на разборчивость речи значительное влияние оказывает переменная составляющая случайной задержки речевого сигнала при передаче по сети связи. Например, доля речевых пакетов, задержка которых превышает на 50 мс, допустимую, не должна превышать 1% от общего количества переданных пакетов [18]. Требуемый уровень изохронности, который может быть допущен в ИТС с пакетной коммутацией, важен по двум причинам [14]. Во-первых, в таких сетях величина переменной составляющей сетевой задержки должна быть, по крайней мере, меньше, чем величина требуемой изохронности передачи. Выбор указанного ограничения на передачу определяется тем, что, например, для передачи речи в силу психофизиологических особенностей человека она должна заканчиваться ко времени возобновления звучания в пункте назначения вновь прибывших речевых сегментов и полезно знать точность, с которой это возобновление звуковых сегментов должно происходить. Во-вторых, проектируемые пакетные ИТС должны обеспечивать поддержание переменной задержки в заданных границах для различных типов изохронного трафика и эти границы должны быть известны. Например, потеря почти половины речевых фрагментов с незначительной длительностью звучания (около 19мс) снижает разборчивость речи лишь на 20%. При этом для фрагментов с длительностью звучания до 250 мс при удовлетворительном воспроизведении речи вероятность потери не должна превышать 1%.

3.4.2. Основные требования, предъявляемые к передаче различных типов информации.

Приведем основные требования к переносу основных типов трафика в рамках предоставления Triple Play услуги.

3.4.2.1. Цифровая речь

Речь традиционно трактуется как трафик от непрерывного источника, имеющий чередующиеся периоды активности и молчания. Речевая информация обладает смысловой избыточностью и допускает значительный уровень искажений. Основное влияние на качество субъективного восприятия речевых сигналов, прошедших через цифровую систему связи, оказывают следующие два фактора:

- длина речевого сегмента (РС);
- сохранение непрерывности РС.

Под речевым сегментом понимается отрезок речевого сигнала, состоящего из системы активных звуковых фрагментов, чередующихся с паузами, длина которых меньше или равна наперед заданной величины.

Речевой сегмент преобразуется и кодируется в цифровой системе как единое целое, образуя своеобразное речевое «сообщение». Временная дискретизация осуществляется на уровне законченных РС, т. е. на уровне непрерывных смыслоразличительных звуков или фонем. Длина РС не настолько коротка, чтобы содержать только одно слово или слог, но и не настолько длинна, чтобы содержать одну фразу или изречение. Обычно длина РС выбирается таким образом, чтобы при его обработке и транспортировке сохранялась непрерывность звучания звуковых фрагментов на приемном конце линейного цифрового тракта (ЛЦТ). Таким образом, при построении любой сети с интеграцией служб необходимо так спроектировать ее систему доставки информации, чтобы существенно уменьшить воздействие переменной (случайной) составляющей сетевой задержки на качество воспроизведения звуковых сигналов.

Известны два основных подхода к обработке и транспортировке речевых сигналов через цифровую систему передачи:

- 1) способ передачи с промежуточным накоплением (пакетизированная речь) или способ передачи с асинхронным временным уплотнением (АВУ);
- 2) способ передачи с синхронным временным уплотнением (СВУ).

За счет появления случайных сетевых задержек речевых пакетов после их

транспортировки по сети нарушает непрерывность звучания на приемном конце. Доля речевых пакетов, задержка которых превышает на 50 мс допустимую, не должна превышать 1% от общего количества переданных пакетов.

Как показывают эксперименты, влияние переменной составляющей сетевой задержки наиболее ощутимо при наличии коротких пауз между словами по сравнению с более длинными паузами между фразами и изречениями. Отсюда следует, что выбор длины РС в сетях с интеграцией служб, в которых реализован способ передачи информации с промежуточным накоплением, необходимо производить с учетом статистики речевых сигналов [18].

Кроме воздействия переменной (случайной) составляющей сетевой задержки, на качество воспроизведения звуковых сигналов у абонента-получателя также влияет и фиксированная (постоянная) задержка, связанная с проводкой и обработкой речевых пакетов по сети и восстановлением из них РС. Задержка передачи речевого сигнала вызывает два нежелательных явления: «эхо» и «наложение речи». Эффект «эхо» (прослушивание абонентом собственного голоса) возникает при величине задержки распространения звукового сигнала от источника к приемнику и обратно больше 50 мс. Согласно рекомендации Сектора по стандартизации телекоммуникаций Международного союза электросвязи (ITU-T) G.114 «наложение речи» (эффект, при котором речь одного говорящего прослушивается в телефоне другого в тот момент, когда он ведет активный разговор) возникает, когда величина односторонней задержки в канале связи составляет более 150 мс [19].

Таким образом, особенности передачи речевых сигналов вытекают из анализа их свойств, а именно:

- определяет заданную величину изохронности передачи речевого сигнала;
- на разборчивость речи значительное влияние оказывает переменная составляющая случайной задержки речевого сигнала при передаче его по сети связи. Доля речевых пакетов, задержка которых превышает на 50 мс допустимую величину переменной составляющей сетевой задержки, не должна превышать 1% от общего количества переданных пакетов;
- при поддержании изохронности передачи особое внимание должно быть уделено сохранению непрерывности передачи речевого сигнала в реальном времени, т. е. обеспечению постоянной составляющей сетевой задержки. Задержка в передаче сигналов от абонента до абонента не должна превышать величины порядка 300 мс в связи с психофизиологическими особенностями восприятия речевых сигналов человеком;
- вероятность ошибки в канале связи допускается порядка 10^{-2} , так как передача речевых сигналов в цифровой форме, как правило, не нуждается в защите от ошибок. При использовании низкоскоростных РПУ возможна защита помехоустойчивым кодом отдельных фрагментов речевых блоков в режиме исправления ошибок;
- сравнительно редкие вставки и/или выпадения фрагментов речевых сигналов слабо влияют на качество воспроизведения речи. В зависимости от избыточности цифрового сигнала на выходе речепреобразующих устройств (РПУ) и длительности речевых сегментов (от 250 до 16 мс), допустимый уровень потерь составляет от 1 до 50% времени активности абонента;
- для повышения использования пропускной способности канала связи при передаче речевой информации необходимо учитывать статистику речевых сигналов, так как доля пауз в телефонном разговоре составляет в среднем 56% в сеансе связи, при этом суммарная продолжительность кратких перерывов в слитной речи длительностью от 5 до 200 мс как внутри слов, так и между словами, занимает около 15%;
- для речевого канала требуется постоянная полоса пропускания шириной от 5,3 Кбит/с до 64 Кбит/с в зависимости от применяемых методов кодирования речевого сигнала для устранения его избыточности.

3.4.2.2. Видеоинформация

Видеоинформация состоит из последовательности неподвижных изображений (кадров), которые воспроизводятся с частотой 25-30 кадр/с. При передаче каждый кадр интерпретируется как объединение некоторого числа фиксированных точек (пикселей) с определенной яркостью и цветом.

При этом величина постоянной составляющей сетевой задержки пакетов для видеоинформации может варьироваться в широком диапазоне: в то время как низкоскоростная 64Кбит/с видеоконференция может допускать величину транзитной задержки порядка 300мс, - высокоскоростная видеоконференция 1,5Мбит/с требует гарантии запаздывания не более 5мс, а для видео HDTV должна быть гарантирована величина равная 1мс. Для потока MPEG-2 указанная величина задержки не должна превышать 4мс (ограниченную 150мс на коммутатор) [16].

Для качественного восприятия плавности движущегося изображения, которое определяется количеством отличающихся изображений в секунду (не менее 25 кадров/с), величина переменной составляющей их сетевой задержки также должна быть, по крайней мере, меньше, чем величина требуемой изохронности передачи. В то время как мерцание зависит только от частоты перерисовки экрана на приеме и может обеспечиваться высокой скоростью сканирования изображений, находящихся в памяти приемника цифрового видео (монитора) самим приемником, например, с частотой 75 и более кадров/с (задержка появления/исчезновения видеоизображений должна заканчиваться до его угасания на сетчатке, где оно остается несколько миллисекунд).

Конкретная скорость передачи зависит от разрешения (числа пикселей в кадре), частоты кадров (числа кадров в секунду), количества информации для представления каждого пикселя (бит на пиксель) и в зависимости от требуемого качества изображения может изменяться в широких пределах: от единиц до сотен мегабит в секунду. Например, для производственных условий, где изображения должны передаваться в реальном времени с постоянной скоростью, для достижения высокого качества видеосигнала, может потребоваться скорость передачи до 1,8 Гбит/с. Указанная пропускная способность необходима также для цифрового кино. Как и речевая информация, видеоинформация также обладает довольно большой избыточностью и при ее передаче могут также применяться различные методы сжатия. В этом случае передача сигнала осуществляется с переменной скоростью, хотя отдельные видеок кадры генерируются с постоянной скоростью, т. е. объем данных для представления каждого кадра различен. Выбор стандарта сжатия (например, MPEG, M-JPEG, H.261) определяет степень сжатия видеоинформации и соответственно качество передаваемого сигнала, а также необходимую полосу пропускания, например, 4-5 Мбит/с (MPEG-2) или 15-21 Мбит/с в случае применения стандарта M-JPEG. Технология MPEG-2 при практически незаметном ухудшении качества позволяет уменьшить скорость оцифрованного несжатого видео с 270Мбит/с до 16 Мбит/с для видео студийного уровня и до 4-5 Мбит/с для видео общего пользования. Стандарт M-JPEG требует полосы 15-21 Мбит/с. Ширина полосы пропускания для передачи компьютерной анимации может варьироваться в широких пределах: от 14,4 кбит/с для анимации на странице Web со сменой кадра в три секунды до потока в несколько Гбит/с.

Существуют два основных типа видеоприложений: интерактивное видео (например, видеоконференции) и потоковое видео (IPTV, которое может использовать как одно-, так и многоадресную рассылку). На основании проведенного анализа рекомендаций МСЭ-Т и IETF обобщим основные требования к характеристикам QoS при реализации передачи видеоданных.

Требования для трафика интерактивного видео

Для интерактивного видео (видеоконференций) к характеристикам QoS предъявляются следующие требования:

- интерактивный видеотрафик (в соответствии с «Базовыми основами QoS») должен быть промаркирован AF41;
- потери – не более 1%;
- однонаправленная задержка – не более 150 мс;
- флуктуация задержки – не более 30 мс;
- минимально-гарантированная полоса пропускания (LLQ) должна быть равна размеру сессии видеоконференции плюс 20%. Например, сессия видеоконференции в 384 кбит/с требует настройки 460 кбит/с полосы трафика гарантированного приоритета.

Так как видеоконференция включает аудиокодек G.711 для речи, то она имеет и соответствующие голосовому трафику требования к потерям, задержке и колебаниям задержки. Однако трафик видеоконференции радикально отличается от трафика голоса.

Требования для трафика потокового видео

Потоковый трафик порождают услуги «видео и аудио по запросу» и Web-вещание. Режим *предоставления аудио- или видеоданных по запросу* предполагает непрерывный доступ к уже готовым файлам мультимедиа на потоковых серверах, которые перед воспроизведением на приеме буферизуются на плеере пользователя. Режим непосредственной трансляции медиапотока на заданную аудиторию с Web-камеры и микрофона называется Web-вещанием или технологией Webcast. Наиболее часто применяемый режим потокового мультимедиа — это передача одноадресного потока. Групповое вещание — альтернативный режим вещания, при котором один медиапоток обеспечивает информацией одновременно многих пользователей данной услуги и требует значительно меньшей полосы пропускания, чем адресное.

Потоковый трафик предъявляет высокие требования к потерям, флуктуации задержки и менее чувствителен к постоянной составляющей задержки. Для потокового видео к характеристикам QoS предъявляются следующие требования:

- потоковое видео (одноадресной или многоадресной рассылки) в соответствии с «Базовыми основами QoS» должно быть промаркировано CS4;
- потери – менее 2 %;
- постоянная составляющая сетевой задержки – менее 4-5 с (в зависимости от возможности буферизации видеоприложений);
- отсутствие значительных требований к флуктуации задержки;
- требования по гарантиям полосы (CBWFO) должны зависеть от формата кодирования скорости видеопотока;
- потоковое видео обычно однонаправленное и поэтому в удаленных филиалах маршрутизаторы можно не настраивать на поддержку потокового видео в направлении от филиала к центру;
- «неважные» приложения потокового видео, такие как видео для развлечения, могут быть промаркированы DSCP CS1 и для них необходим минимум гарантий полосы пропускания в очереди CBWFO (с использованием класса Интернет/scavenger).

Для мониторинга качества при передаче видеопотоков вводится новая метрика – QoE (Quality of Experience). Для IPTV QoE определяется в проекте рекомендации G.IPTV – QoE. Метрика QoE используется не только для IPTV, но и для, например, аудиоинформации. Метрика QoE регламентируется в рекомендации P10/G100 и в соответствии с этой рекомендацией определяется как глобальная приемлемость приложений или услуг, субъективно воспринимаемая конечным пользователем. При этом отмечается, что экспериментальное качество включает в себя предоставляемое качество всеми элементами, участвующими в предоставлении и получении услуг из конца в конец, - терминалами, сетями, инфраструктурой услуг и т.д., а также качество восприятия услуг клиентом. Кроме того, глобальная приемлемость может зависеть от конкретно предоставляемого контента и расположения пользователя к услуге.

3.4.2.3. Асинхронный трафик (класс C)

Асинхронный трафик, в отличие от изохронного, допускает сравнительно большие вариации постоянной задержки (определяемые прикладными применениями) и не критичен к поддержанию изохронности при передаче по каналам связи. Однако предъявляет достаточно жесткие требования к достоверности передачи (порядка 10^{-5} – 10^{-7} на бит) и к сохранности информации (вероятность засылки не по адресу порядка 10^{-6} на пакет), так как утраченные данные восстанавливаются за счет повторной передачи. Пользователям чаще всего требуется независимый темп передачи и приема данных, многорежимный обмен (интерактивный обмен данными, передача файлов), обеспечение конфиденциальности. Не допускаются вставки и/или потери отдельных элементов потока данных. Очень важным требованием во многих применениях является сохранение порядка следования данных. Интенсивность посылки пакетов асинхронного трафика в сеть и их размер могут изменяться в широких пределах, например, коэффициент пульсаций трафика (отношения максимальной мгновенной интенсивности трафика к его средней интенсивности) протоколов без установления соединений может достигать до 200, а протоколов с установлением соединений – до 20. Требования к ширине полосы пропускания асинхронного трафика лежат в широких диапазонах: от десятков кбит/с для низкоскоростных интерактивных приложений до сотен Мбит/с для приложений, ориентированных на работу с графическими данными. В таблице 3.2. показаны допустимые значения потери пакетов для различных информационных потоков [20].

Таблица 3.2. - Допустимые значения вероятности потери пакетов

Трафик	Формат	Допустимые значения
Речь обычного качества	МККТТ G.711. ИКМ (64 кбит/с)	$<10^{-3}$
Речь высокого качества	МККТТ G.727. Полосная АДИКМ (64 кбит/с)	$<10^{-5}$
Телевидение обычного качества	Сжатие сигналов (средняя скорость 10 Мбит/с)	$<10^{-9}$
Телевидение высокой четкости	Сжатие сигналов (средняя скорость 100 Мбит/с)	$<10^{-10}$
Передача данных	HDLC (от 64 кбит/с до 100 Мбит/с)	$<10^{-6}$

3.4.2.4. Мультимедиа

Требования, накладываемые на ИТС сводятся к двум аспектам: требования к качеству передачи трафика и функциональные требования: организация мультикастинга; безопасность (аутентификация, целостность, конфиденциальность); управление сессиями (описание типа мультимедиа, оповещение о сессии, идентификация сессии, синхронизация мультимедиа) и др. При этом требования к передаче мультимедиа-объектов определяются комбинацией различных видов трафика, передаваемых по сети. Кроме того, при передаче мультимедийного потока для устранения смещения (skew) по времени может потребоваться межпоточковая синхронизация изохронных потоков, так как, например, для обеспечения синхронизации речи с движением губ на приеме skew между аудио- и видеоинформацией не должно превышать 120мс [18]. Проблема межпоточковой синхронизации является одной из составных частей проблемы обеспечения QoS-норм переноса приложений мультимедиа. При этом процедуры механизмов защиты информации, например, аутентификации и авторизации пользователей ИКС необходимо включить в интерактивные сценарии их взаимодействия с контентом Web-, видео- или аудио- приложений в процессе установления соединения, о чем речь пойдет ниже.

Противоречивость требований к качеству передачи информации обуславливает необходимость создание ИКС с таким набором сетевых служб, чтобы обеспечить возможность доставки всего пакета инфоуслуг с заданным сквозным качеством QoS на базе единой сетевой инфраструктуры со специальными комбинированными процедурами обслуживания и дообслуживания очередей. При этом процедуры механизмов защиты информации, например, аутентификации и авторизации пользователей ИТКС необходимо включить в интерактивные сценарии их взаимодействия с контентом Web-, видео- или аудио- приложений в процессе установления соединения, о чем речь пойдет ниже. Противоречивость требований к качеству передачи информации обуславливает необходимость создание ИКС с таким набором сетевых служб, чтобы обеспечить возможность доставки всего пакета инфоуслуг с заданным сквозным качеством QoS на базе единой сетевой инфраструктуры со специальными комбинированными процедурами обслуживания и дообслуживания очередей.

3.5. Особенности передачи речевых сигналов в пакетных ИТС. Обеспечение QoS-норм в пакетной и гибридной транспортной системе

Одной из главных проблем в ИТС также остается проблема обеспечения изохронности переноса информационных потоков класса B и их синхронизации при транспортировке мультимедийных объектов в рамках единой универсальной услуги связи. Обеспечить изохронность переноса порций информации в сети можно двумя способами:

1) либо строго фиксировать постоянное время доставки каждой порции информации через ИТС;

2) либо так ограничить трафик на ее входе, чтобы интервалы времени, соответствующие фиксированному значению квантиля распределения времени задержки порций информации в системе, не превышали заданной величины.

При первом способе должны быть полностью исключены случайные задержки информации в транспортной системе. Это возможно лишь в том случае, если ее физические ресурсы в сессии предоставляются каждой порции информации без ограничений и конфликтов доступа, т. е. со статическим или «жестким» закреплением. Необходимо отметить, что физические ресурсы являются компонентами общесистемных ресурсов, в качестве которых выступают логические каналы соответствующих уровней архитектуры ИТС (h -ресурсы и h -каналы в соответствии с терминологией ВОС) и бесконфликтное закрепление общесистемных ресурсов более высоких уровней ИТС не влечет за собой бесконфликтного закрепления общесистемных ресурсов более низких уровней. Например, несколько транспортных каналов транспортного уровня могут использовать один виртуальный канал сетевого уровня, или несколько сетевых виртуальных каналов могут использовать один и тот же логический канал уровня звена. При этом система покласовых приоритетов при использовании общесистемных ресурсов для реализации конкретного транспортного соединения не приводит к статическому закреплению физических ресурсов ИТС, т. е. даже для информации высшего приоритета перед общесистемным ресурсом может возникнуть очередь из порций информации данного приоритета, принадлежащих разным соединениям. Отсюда следует, что неограниченное бесконфликтное использование физических ресурсов в ИТС возможно только при индивидуальном их закреплении (а следовательно, и индивидуальном закреплением одновременно всех общесистемных ресурсов системы) за B -соединением между двумя сеансовыми B -объектами в мультимедийной сессии. Вполне естественно, что в соответствующем типе ИТС должен существовать механизм такого сквозного бесконфликтного закрепления физических ресурсов за B -соединением в рамках мультимедийного соединения.

При втором способе допускаются случайные задержки информации в ИТС, которые не превосходят некоторой фиксированной величины, т. е. осуществляется ограниченный доступ к ее физическим ресурсам на коллективной основе (определяющий признак

пакетных ТС). При этом должны быть приняты некоторые специальные меры, обеспечивающие допустимые фиксированные задержки для отдельных порций информации в сети.

3.6. Функционально-структурная организация инфотелекоммуникационных транспортных систем на технологиях IP-QoS

Можно утверждать, что широкополосные высокоскоростные МСС будущего и будут строиться на базе пакетной коммутации. Это связано с более экономичным использованием сетевых ресурсов.

В России для построения региональных (зоновых) компонентов национальной МСС рекомендованы две базовые пакетные технологии IP-QoS (технология интегральных услуг с резервированием ресурсов (Integrated Services, IntServ), технология дифференциальных услуг (Differentiated Services, DiffServ) в сочетании с технологией многопротокольной коммутации по метке (Multi-Protocol Label Switching, MPLS)) и технология ATM – для построения ее магистральной компоненты. Технология ATM была рекомендована также для внедрения мультимедийных услуг в нашей стране и Постановлением Минсвязи России № 176 от 29 января 1997г.

3.6.1. Специфика функционально-структурной организации инфотелекоммуникационных транспортных IP-QoS-систем

Основным сдерживающим фактором создания ИКС на технологии IP является проблема обеспечения QoS-норм для трафика различной природы (особенно для изохронных приложений реального времени). Необходимость арбитража и адаптивного перераспределения сетевых ресурсов потребовала привнесения в телекоммуникационную инфраструктуру интеллекта на базе критериев QoS. В среде IP существует два механизма:

Абсолютный механизм QoS, резервирующий ресурсы узлов сетевой инфраструктуры из конца в конец и гарантирующий соответствие параметров трафика. Каждый информационный поток нуждается в резервировании ресурсов, гарантирующих качество его транспортировки. Информационный поток эквивалентен пяти типам идентификаторов: транспортный протокол, адрес и порт источника, адрес и порт получателя.

Релевантный (относительный или сравнительный) механизм QoS, основанный на структурировании информационных потоков с помощью приоритетов, позволяющих агрегировать информационные потоки с близкими требованиями к качеству в ограниченный набор классов (CoS), которые делят сетевые ресурсы между собой в соответствии с назначенным им приоритетом. При этом каждый пакет имеет поле принадлежности к тому или иному классу, которое содержит признак алгоритма прохождения через узлы магистральной сети трафика данного типа.

Для решения этой проблемы группой IETF на базе протокола IP были предложены две модернизированных технологии: технология интегральных услуг с резервированием ресурсов по протоколу RSVP (Integrated Services/RSVP, IntServ) и технология дифференциальных услуг с предварительной организацией соглашения об уровне сервиса SLA (Differentiated Services/SLA, DiffServ).

Технология IntServ [RFC1633] реализует абсолютный механизм QoS и резервирует для каждого соединения сетевые ресурсы от приемника в симплексном режиме. Для явного задания уровня QoS предполагает наличие в маршрутизаторах сети четырех основных компонентов: программы–демона RSVP [RFC2205], механизмов контроля допустимостью соединения (Connection Admission Control, CAC), а также классификатора пакетов (Packet Classifier) и планировщика пакетов (Packet Scheduler). Модель IntServ может обеспечить сквозное (end-to-end) качество обслуживания, гарантируя необходимый уровень полосы пропускания, если все маршрутизаторы в сети поддерживают и реально

учитывают механизмы IntServ. Конечный узел инициирует QoS-сеанс IntServ, посылая запрос RSVP — сигнального протокола, служащего для резервирования ресурсов в сети. Сеть может ответить или согласием, или отказом — все зависит от доступности ресурсов на каждом ее участке. При «положительном решении вопроса» полоса пропускания резервируется. Каждый маршрутизатор в сети поддерживает таблицу состояний (state table) для каждого сеанса IntServ. В модели IntServ состояние резервирования необходимо поддерживать во всей сети, что накладывает дополнительную нагрузку на ЦПУ и ОЗУ сетевых устройств. Причем каждое устройство на пути следования пакета, включая конечные узлы, должно «понимать» алгоритмы IntServ.

Процесс резервирования ресурсов соответствующего узла: специальная программа — демон RSVP — обращается к двум модулям принятия решения: модулю управления доступом и административному. Управление доступом необходимо для определения наличия на соответствующем узле необходимых для обеспечения QoS ресурсов. Административный модуль должен проверить наличие прав на резервирование у конечного узла. При отрицательном исходе любой из проверок RSVP возвращает на удаленный запрашивающий узел уведомление об ошибке, означающее отказ в маршрутизации данного информационного потока. Если обе проверки дали положительный ответ, то RSVP устанавливает в соответствии с запросом параметры гарантии качества в классификаторе пакетов — модуле, различающем «окраску» трафика, а также в планировщике пакетов, определяющем порядок следования пакетов, обеспечивающий заданную полосу пропускания.

Одно из основных достоинств RSVP — это масштабируемость до очень больших групп с многоадресной доставкой пакетов. Он использует ориентированные на приемный узел запросы, которые сливаются по мере их прохождения по многоадресному дереву. При этом запрос на резервирование посылает не узел, собирающийся вести многоадресную трансляцию, а каждый из узлов-получателей. Такая схема позволяет отказаться от следования сигнала резервирования из конца в конец сети и ограничить его маршрут ближайшим узлом, в котором сливаются подобные сигналы других ветвей.

Таким образом, функцию «управление резервированием сетевых ресурсов» в IntServ реализует сигнальный протокол RSVP. Сигнальный протокол RSVP работает поверх IPv4 или IPv6. Сообщения RSVP посылаются от узла к узлу между RSVP-маршрутизаторами, поддерживающими этот протокол, в виде IP-дейтограмм с кодом протокола 46. Запрос резервирования инициирует формирование мультимедийной сессии с заданными параметрами и включает в себя набор опций, которые в совокупности называются стилем. RSVP обеспечивает несколько моделей резервирования (стилей), для того чтобы удовлетворить требованиям различных мультимедийных приложений. Они определяют способ резервирования различными отправителями в пределах одной сессии и контроль выбора отправителей. В настоящее время определены следующие стили: WF (Wildcard-Filter), FF (Fixed-Filter) и SE (Shared Explicit). Обобщая, можно сказать, что RSVP имеет следующие атрибуты:

- RSVP выполняет резервирование для уникастных и мультикастных приложений, динамически адаптируясь к изменениям членства в группе вдоль маршрута. Для уникастных приложений требуемое QoS будет получено с помощью диспетчера пакетов в драйвере уровня звена данных. При групповом резервировании учитывается тот факт, что в точках слияния дерева доставки несколько потоков, для которых производится резервирование, сливаются в один, так что подтверждающее сообщение передает маршрутизатор, находящийся в точке их слияния.

- RSVP является симплексным протоколом, т. е., он выполняет резервирование для однонаправленного потока данных. RSVP ориентирован на получателя, т. е., получатель данных инициирует и поддерживает резервирование ресурсов для потока. RSVP

поддерживает динамическое членство в группе и автоматически адаптируется к изменениям маршрутов.

- RSVP не является маршрутным протоколом, но зависит от существующих и будущих маршрутных протоколов.
- RSVP транспортирует и поддерживает параметры управления трафиком и политикой, которые остаются непрозрачными для RSVP.
- RSVP обеспечивает несколько моделей резервирования или стилей, для того чтобы удовлетворить требованиям различных приложений.
- RSVP обеспечивает прозрачность операций для маршрутизаторов, которые его не поддерживают.
- RSVP может работать с IPv4 и IPv6.

Очевидно, что протокол RSVP наиболее эффективен для многоадресных приложений, то есть систем группового вещания (video broadcasting или video near on demand VNOD) в датаграммных сетях. В то время как для VOD-или VoIP-приложений он аналогичен по сложности прокладке виртуального соединения в ATM.

Функциональная структура маршрутизатора и оконечной системы сети IntServ показана на рис.3.6.

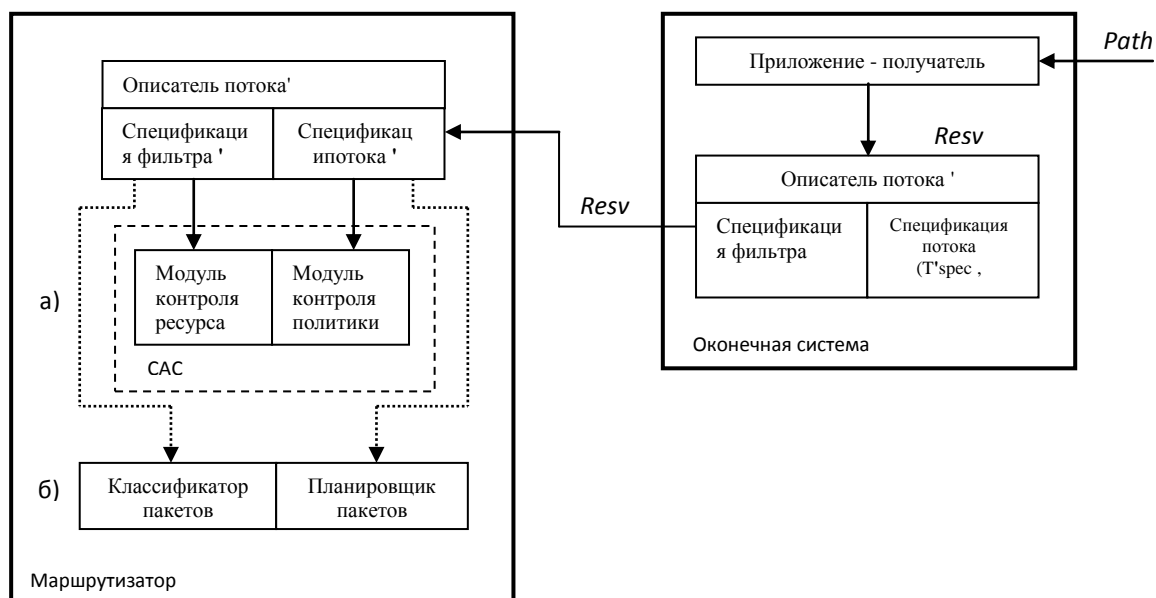


Рис. 3.6. Функциональная структура маршрутизатора и оконечной системы сети IntServ

При получении сигнального запроса *Path* получатель отправляет соответствующее подтверждение источнику в виде сообщения «резервирование» *Resv* по проложенному им маршруту, которое сообщает транзитным маршрутизаторам параметры для резервирования сетевых ресурсов, принимаемые приемником в сеансе связи с источником. Эти параметры содержатся в дескрипторе потока TD (Traffic Descriptor). В процессе резервирования сетевых ресурсов используется оперативная информация о каждом потоке, которая содержится в дескрипторе потока, включающего спецификации фильтра *filter spec* и потока *flowspec*. В технологии IntServ функция CAC реализуется в маршрутизаторах посредством двух модулей: модуля «контроль доступа» и административного модуля «управление политикой». Если запрос поддержан, то в каждом маршрутизаторе резервируются буферное пространство и полоса пропускания, а в классификаторе и планировщике пакетов, в соответствии с запросом параметров QoS, осуществляется параметризация потока в целях его классификации и управления в сессии. Механизм обеспечения QoS включает в себя классификацию пакетов, административный контроль и диспетчеризацию.

В технологии IntServ дополнительно к услуге «наилучшей попытки» BE реализованы две услуги: контролируемой загрузки (Controlled Load Service, CLS) и гарантированной доставки (Guaranteed Service, GS). *Услуга контролируемой загрузки CLS [RFC2211]* в загруженной сети обеспечивает качество передачи выше, чем может предоставить услуга «наилучшая попытка» BE, однако какие-либо гарантии по переносу трафика при этом также не предоставляются. Пакеты получают максимум из возможного в загруженной сети, но без жестких гарантий QoS. Один из возможных подходов реализации услуги CLS – это реализация отдельной приоритетной буферизации пакетов различных услуг. Один из возможных подходов реализации услуги контролируемой загрузки CLS в реальном оборудовании IntServ заключается в реализации отдельной приоритетной буферизации пакетов различных услуг является использование алгоритмов PQ. *Услуга гарантированной доставки GS [RFC2212]* предоставляет гарантии только по полосе пропускания и максимальной задержке пакетов «из конца – в конец» сети. Под гарантиями по полосе пропускания подразумевается, – что при соблюдении потоком продекларированных параметров не произойдет потерь пакетов в очередях, а под гарантиями по задержке, – что максимальная задержка «из конца–в конец» не превысит заранее определенной верхней границы. Для реализации услуги гарантированной доставки GS используют алгоритм WFQ. Таким образом, несмотря на расширение набора функций гарантии по дисперсии задержки пакетов в сети технология IntServ не обеспечивает. Кроме того, архитектура IntServ плохо масштабируема: увеличение числа обслуживаемых соединений ведет к снижению производительности маршрутизаторов. Это предопределило разработку более простого протокола DiffServ, лишённого указанных выше недостатков.

Использование IntServ/RSVP может обеспечить гарантии абсолютной схемы QoS, при которой, для обеспечения абсолютных гарантий, ресурсы узлов сети резервируются из конца в конец сети. ***Однако, некоторая слабость RSVP, обусловленная необходимостью обработки пакетов в каждом узле маршрута, предшествующей установлению соединения, сдерживает его широкое распространение.*** Обработка включает:

- сигнализацию из конца в конец сети;
- идентификацию информационных потоков;
- трассировку и гарантии резервирования ресурсов;
- политику управления трафиком при резервировании пропускной способности узлов;
- планирование транспортировки трафика в соответствии с требуемым качеством.

Это особенно важно при большом количестве мультимедийных информационных потоков, которое достигает у некоторых сервис-провайдеров США, использующих каналы OC-192, нескольких миллионов.

Многие приложения нуждаются в QoS, но не имеют возможности получить необходимые гарантии с помощью IntServ. ***До 50% транзакций Internet имеют длительность всего несколько секунд, что приводит к тому, что процесс резервирования ресурсов для них существенно превышает приемлемое время ожидания.***

Модель дифференцированного обслуживания (Differentiated Services — DiffServ, RFC2475) решает некоторые проблемы ToS и IntServ. Она более масштабируемая, и при корректной реализации может работать в инфраструктуре, охватывающей несколько сетей. Первые шесть битов ToS в пакете IPv4 или октет “класс трафика” в пакете IPv6 называются полем DiffServ CodePoint (DSCP), в котором можно указывать до 64 классов трафика.

Сеть маршрутизаторов с поддержкой механизмов DiffServ называют “облаком DiffServ”. Классификация трафика происходит на входе в это “облако”. Провайдер

обычно договаривается с клиентом и заключает с ним соглашение об уровне обслуживания (SLA). Например, компания может подписаться на бронзовый, серебряный или золотой пакет услуг Интернет-провайдера. Заключенный контракт и определит уровень приоритетов DiffServ.

Качество краткосрочных транзакций Internet более эффективно обеспечивается относительной схемой QoS.

DiffServ Differentiated Services (DiffServ или просто DS) **обеспечивает простой, но грубый, метод структурирования трафика по классам сервиса CoS с соответствующими приоритетами.** Он использует поле типа сервиса в заголовке IP-пакета, называемое DS byte, для определения алгоритма его транспортировки от узла к узлу сети. Такая схема, в отличие от IntServ, не требует затрат времени на предустановку параметров транспортировки трафика по маршруту следования информационного потока.

Согласно датаграммным принципам построения сети, DiffServ транспортирует информацию QoS в каждом пакете, агрегируя все информационные потоки в ограниченное число (максимум 64) классов. **Пока стандартизовано только число классов, агрегирующих информационные потоки в группы, различающиеся по алгоритмам транспортировки трафика per hop behavior (PHB).**

Вопросы расщепления трафика пока не определены. Помимо алгоритма BE, DiffServ предусматривает еще два варианта алгоритмов: ускоренная транспортировка Expedited Forwarding (EF), уверенная доставка Assured forwarding (AF).

Ускоренная транспортировка EF предусматривает малые потери, низкую латентность и малый джиттер для информационных потоков этой группы. Этот алгоритм применим для приложений Internet-телефонии, видеоконференций или услуг виртуальных выделенных каналов. Уверенная доставка AF объединяет абонентов, нуждающихся в предсказуемых услугах с постоянной скоростью доставки. Степень предсказуемости должна быть определена в каждом конкретном случае. В рамках каждого класса AF, могут быть предусмотрены три уровня процедур предотвращения переполнений.

Принадлежность к тому или иному классу обслуживания определяется значением DiffServ code point (DSCP), содержащимся в заголовке пакета. В каждом маршрутизаторе пакеты обрабатываются в соответствии с алгоритмом транспортировки, определяемым классом CoS. На сегодняшний день стандартизовано 14 DSCP.

В узлах коммутации/маршрутизации информационных потоков логично каждый класс представить своей очередью, механизмы работы, которых аналогичны применяемым в IntServ (WFQ, WRED и т.п.).

Основное достоинство DiffServ — это его масштабируемость, так как отсутствует необходимость устанавливать параметры транспортировки для каждого из информационных потоков, группируемых в классы таким образом, чтобы суммарная информация о состоянии узлов сети объединяла разрозненные описания пропорционально числу классов.

DiffServ не способен маршрутизировать пакеты, просто игнорируя значение DSCP в их заголовках, и обрабатывать пакеты по алгоритму BE.

Самым большим преимуществом модели DiffServ является то, что она действует на границе “облака”. После того как данные пересекли эту границу, внутренним маршрутизаторам можно не заниматься поддержанием информации о статусе QoS и полностью сосредоточиться на своей основной функции — маршрутизации.

Но и при использовании модели DiffServ остаются элементы непредсказуемости. Отдельные внутренние маршрутизаторы могут неадекватно отреагировать на значения битов в поле ToS или даже изменить их. И нет

установленных стандартов: “золотой” статус одного провайдера может соответствовать “бронзовому” другого. Следовательно, вы можете платить своему Интернет-провайдеру за самое лучшее обслуживание, а при переходе в сеть другого провайдера статус вашего трафика изменится. Поскольку DiffServ работает за счет выборочного сброса пакетов в периоды сетевой перегрузки, то соединения с низким приоритетом вообще могут разорваться во время “всплесков” сетевой активности.

Модель DiffServ хороша для крупных ЛВС и территориально распределенных сетей (WAN), так как имеет меньшие накладные расходы (overhead) и лучше масштабируется (по сравнению с IntServ). Поскольку классификация трафика происходит на входе в “облако” DiffServ, то конечные узлы и промежуточные маршрутизаторы не обязаны ни “понимать”, ни устанавливать значения битов DiffServ.

В технология DiffServ [RFC2475] реализован релевантный (относительный или сравнительный) механизм QoS, позволяющий выделять для агрегированных потоков с близкими требованиями к QoS-нормам ограниченный набор классов сервиса (CoS), которые делят сетевые ресурсы между собой в соответствии с назначенным им приоритетом. Технология DiffServ [RFC2475] ориентирована на источник, ответственный за QoS. Высокие масштабируемость и гибкость DiffServ достигаются за счет структурирования потоков с соответствующими приоритетами в несколько агрегированных по уровню сервиса классов CoS.

В сети DiffServ функции организации мультимедийного соединения и *управления резервированием сетевых ресурсов* реализуется либо сигнальным протоколом RSVP+ [RFC3209], либо непосредственно провайдером услуг в процессе заключения с пользователем «трафик-контракта» (Service Level Agreement, SLA) [RFC2475] на поставку определенного класса сервиса обслуживания CoS. Расширение RSVP предполагает некоторое видоизменение стандартной процедуры обработки сообщений RSVP на маршрутизаторах, в частности контроля допуска к агрегированным классам CoS и поддержку агрегированного и туннелируемого RSVP. В технологии DiffServ используется модифицированный протокол RSVP (RSVP+), в котором расширены функции, отвечающие за резервирование ресурсов для агрегированного потока. Определены конкретные типы резервирования ресурсов:

- индивидуальное резервирование (distinct reservation);
- долевое резервирование (shared reservation)
- резервирование полосы для группы пользователей (wildcard filter type reservations),
- долевое явное резервирование (shared explicit type reservation).

Технология DiffServ исключает необходимость учета состояния потока и индивидуальной обработки каждого из них. Предназначенный для использования с DiffServ протокол RSVP+ позволяет приложению идентифицировать себя в сеансе связи и получить соответствующее его приоритету обслуживание. Это расширение RSVP предполагает, в частности, контроль допуска к агрегированным классам DiffServ и поддержку агрегированного и туннелируемого RSVP.

В отличие от RSVP, протокол RSVP+ ориентирован на источник, ответственный за QoS. При установлении соединения RSVP+ при «окраске» потока дополнительно определяет механизмы, которые могут быть задействованы для реализации конкретных услуг в сеансе связи, например, Premium Service, Assured Service и т. д.

Функция *контроля допустимости установления соединений* фактически реализуется путем соответствующих настроек в базе данных граничного маршрутизатора. Резервирование ресурсов и параметризация услуги осуществляется в соответствии с заявленным CoS. Структура граничного маршрутизатора домена DiffServ показана на рис. 3.7.

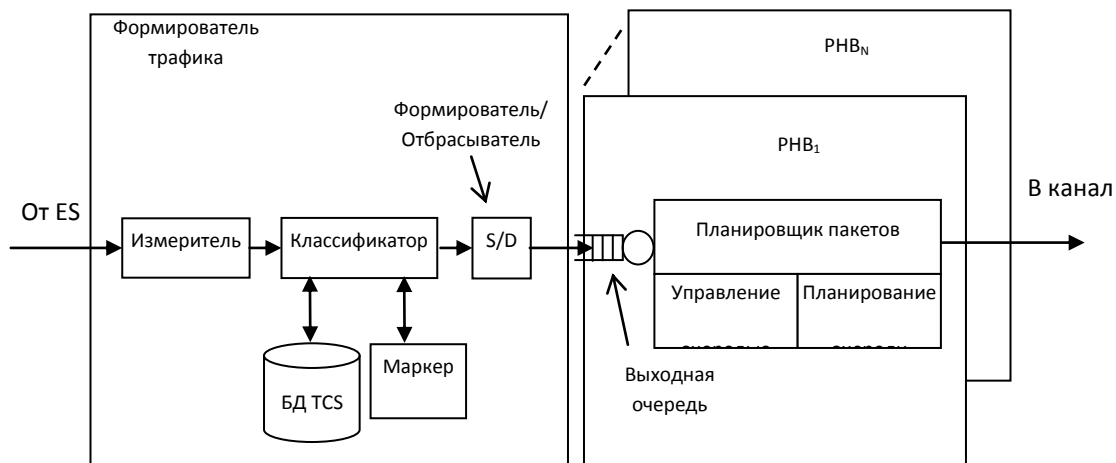


Рис. 3.7 - Структура граничного маршрутизатора домена DiffServ

Концепция обслуживания CoS в сессии здесь обеспечивается тремя механизмами: классификатором пакетов, формирователем трафика TC и механизмом пошаговой маршрутизации PHB. В сессии для каждого соединения осуществляется проверка на соответствие поступающего трафика заявленным параметрам. Эти функции реализованы в формирователе трафика TC, в состав которого входят: маркер (Marker) – формирует код DSCP; классификатор (Classifier) – сортирует пакеты от различных приложений и передает их механизмам управления очередями и планирования; измеритель (Meter), используемый для контроля политики доступа, и формирователь/отбрасыватель (Shaper/Dropper), служащий для формирования профиля входного трафика. Правила работы классификатора описаны в TCS. В соответствии с классом обслуживания CoS функцией маркировки TC пакетам каждого агрегированного потока ставится в соответствие класс пошаговой маршрутизации PHB [RFC2475].

Режим PHB можно рассматривать как совокупность параметров, в соответствии с которыми маршрутизатор устанавливает порядок направления пакетов на интерфейс вывода. Множество пакетов с одинаковым типом PHB называется «общее агрегированное поведение» (foo traffic aggregate, FTA). Внутренние маршрутизаторы домена DiffServ осуществляют лишь перенаправление пакетов по выбранному алгоритму PHB.

В технологии DiffServ реализованы два класса дополнительных услуг: «срочная доставка» EF и «уверенная доставка» AF.

Услуга срочной доставки EF [RFC 2598] (далее PHB EF) предоставляет пользователю гарантии по полосе пропускания и сквозной сетевой задержке для пакетов, аналогичные услуге гарантированной доставки GS в IntServ. Рекомендованное значение поля DS – 101110. При этом требования к вероятности потери пакета, значениям задержки и ее дисперсии, необходимой полосе пропускания и т. д. гарантируются только в рамках домена DiffServ, на узлах которого предоставляется класс услуг PHB EF. Эта услуга может быть востребована для услуги «Premium Service», являющейся аналогом услуги «виртуальная арендованная линия». Услуга срочной доставки PHB EF предоставляет пользователю гарантии по полосе пропускания и сквозной сетевой задержке для пакетов аналогично услуге гарантированной доставки GS в IntServ.

Услуга уверенной доставки AF [RFC 2597] (далее PHB AF) поддерживает уровень качества обслуживания более низкий, чем класс срочной доставки PHB EF, но более высокий, чем обслуживание с BE. Таким образом, услуга PHB AF гарантирует, что высокоприоритетные потоки получают лучшее обслуживание, чем низкоприоритетные. При этом его сквозная задержка в сети может оказаться значительной. Услуга PHB AF позволяет реализовать четыре класса CoS с тремя уровнями приоритета пакета для каждого из них. В маршрутизаторах каждый класс представлен своей очередью. Услуга AF может быть реализована с использованием таких механизмов, как WFQ в сочетании с взвешенным ранним обнаружением переполнений WRED. Механизм WFQ для каждой

очереди назначает вес (приоритет), который определяет, каким образом очереди распределяют полосу пропускания.

Для гарантии качества услуг приложений, использующих информационные потоки реального времени, наиболее важными являются механизмы, реализующие абсолютную схему QoS. Абсолютная схема качества обслуживания QoS в IP-сетях на технологии DiffServ может быть обеспечена двумя современными концепциями, не связанными с природой IP: протоколом многопротокольной коммутации по метке MPLS и протоколом RSVP+.

Новая концепция обеспечения QoS в IP-сетях

Ограничения магистральной сети традиционной IP-среды, связанные с ограниченным числом частных IP-адресов, могут быть преодолены за счет внедрения двух технических возможностей не связанных с природой IP:

Концепция туннелирования, использующая блочную интерпретацию заголовков пакетов, инкапсулированную в заголовки третьего (сетевого) уровня так, что скрывает поля IP-адресации от магистральной сети общего пользования. Эта идея является основой для всех виртуальных приложений, требующих туннелирования частных IP-адресов по магистральной сети общего пользования.

Концепция путей, называемая Label Switched Paths — LSPs, которая использует протоколы маршрутизации для прокладывания путей из конца в конец сети в соответствии с топологическими особенностями маршрутов и доступными сетевыми ресурсами. Распределенная система управления трафиком, управляющая прокладкой пути.

Новая концепция построения IP-сетей основана на ряде взаимодополняющих технологий:

- протокол MPLS;
- расширения RSVP;
- маршрутизация на постоянной основе constraint-based routing (CBR), которые используются для маршрутизации трафика через виртуальные пути статической маршрутизации. Такое решение предотвращает перегрузку маршрутизаторов.

Multiprotocol Label Switching — MPLS. Алгоритмы MPLS не используют специфичные технологические процессы. Например, ATM/FR протокол сигнализации PNNI или транспортировка ATM OAM. ***Протокол MPLS является, фактически, надстройкой над вторым (канальным) уровнем сетевой модели ISO/OSI, призванным обеспечить портирование IP-сервисов третьего (сетевого) уровня модели на различные канальные технологии.*** MPLS сочетает достоинства, с одной стороны, механизмов обеспечения QoS в ATM, построенных по абсолютной схеме QoS, планирующей ресурсы узлов сети, входящих в трассу виртуального соединения, установление которого гарантирует качество обслуживания абонентов, а с другой стороны, поддерживает достоинства IP-среды для маршрутизации коротких транзакций в датаграммных сетях.

Обычно решение о переносе информационного потока в узлах датаграммной сети основано на вычислении очень длинных префиксных выражений в таблицах маршрутизации уникальных информационных потоков (unicast table), адресов назначения, содержащихся в заголовках IP-пакетов. ***MPLS — это схема транспортировки потоков, в которой маршрутизация основана на фиксированных выражениях, содержащихся в MPLS-таблицах «коротких» меток, которые в основном являются частью заголовка MPLS. Маршрут, прокладываемый по сети, для всех пакетов с одинаковыми метками MPLS называется label switched path (LSP).*** Этот путь устанавливается с помощью протоколов RSVP+/LDP или CR-LDP.

На рис. 7 хорошо видно, среды ATM и Frame Relay, ориентированные на установление виртуальных соединений, имеют на своем канальном уровне

механизмы идентификации виртуальных каналов и путей в узлах коммутации, что позволяет им, в отличие от дайтаграммных IP-пакетов, избавиться от избыточных операций по определению направления движения каждого пакета.

Рис. 7. Архитектура MPLS

Дополнение среды IPoEthernet схемой меток (списком узлов, входящих в маршрут информационного потока), позволило привнести аналогичные механизмы и в канальный уровень Ethernet, отказавшись от дайтаграммных алгоритмов маршрутизации. Концептуальная однородность алгоритмов обработки пакетов на канальном уровне для всех, обсуждаемых сред доставки трафика, позволила, за счет алгоритмов автоматического туннелирования, устранить функциональные различия в маршрутизации пакетов и выработать правила преобразований IP-пакетов в пакеты канального уровня. Пакеты канального уровня отличаются от IP-пакетов соответствующими заголовками («конвертами»).

Узел доступа LSP (узел, в котором стартует путь информационного потока) использует фильтры для селекции пакетов, которые посылаются по части путей LSP и фильтры размещают корректные заголовки MPLS в начало пакета, В конечном узле пути LSP (узел, в котором заканчивается путь информационного потока) удаляет заголовок MPLS и пакет маршрутизируется по обычной дайтаграммной схеме. Метка имеет локальное значение. Следовательно, в транзитном узле пути LSP каждая точка входа таблицы транспортировки описывается как набор меток. Эта концепция называется обменом меток и аналогична используемой в ATM схеме транспортировки трафика. **Фактически метка пути очень похожа на идентификатор виртуального соединения ATM, а путь LSP очень похож на виртуальный канал ATM. Различия лишь в размерах и структуре пакетов.**

RSVP+. Изначально, протокол RSVP был создан как инструмент, обеспечивающий транспортировку агрегатов данных без дайтаграммной обработки. Для автоматизации прокладки путей LSPs были определены расширения (новые типы объектов). Наиболее важно использование следующих **расширений**:

Распределение меток для предустановки путей LSP между двумя коммуникационными маршрутизаторами меток LSR.

Спецификация, предопределяющая четкий путь для LSP, не зависит от маршрутизации IP, **Четкий путь — это маршрут, который переносит список узлов специфицируемых оператором.** Этот список содержит заявки на предустановку параметров узлов сети, которые могут быть жесткими или произвольными. Если список содержит только жестко предустановленные узлы, то промежуточные узлы доступны, но не предустановленные узлы должны работать в соответствии со спецификацией.

Поддержка DiffServ. В сетях с поддержкой DiffServ существует два типа LSP: Exp-Label Switch Path E-LSPs и Label-Label Switch Path L-LSPs. E-LSP - это путь, содержащий несколько различных требований QoS LSP, которые обеспечивают до 8 различных классов SLA. Значение DSCP IP пакетов, для входного узла пути LSP, полностью размещается в экспериментальном поле заголовка схемы MPLS. В поле расширения переносятся как алгоритмы обработки запросов, так и процедуры транспортировки пакетов. Аналогично тому, как в сети с поддержкой DiffServ устанавливается соответствие между алгоритмами проноса трафика (PHB) и соответствующим классом трафика, определяемым DSCP, **в DiffServ домене MPLS устанавливается соответствие между алгоритмом проноса трафика PHB и полем расширения, носящим локальный характер**, определяемым сервис-провайдером и конфигурирующим каждый путь LSP. **Коммутируемый путь от метки к метке (L-LSP) описывает путь LSP информационного потока с одним набором параметров QoS.** Соответствие классов QoS определяется метками MPLS,

поэтому возможно определить алгоритм проноса трафика в поле расширения. В этом заключается различие между метками поля расширения и PHB. Пути E-LSPs и L-LSPs могут быть установлены с резервированием полосы пропускания или без него.

Constraint-based routing — CBR или маршрутизация на постоянной основе определяется транспортировкой пакетов внутри сетевого домена, осознанным резервированием ресурсов и концепцией маршрутизации. Она часто ассоциируется с «качеством услуг маршрутизации» («QoS routing»). В основном это алгоритмы определения наикратчайших путей, которые были модифицированы с учетом специфичных ограничений таких как: требования полосы пропускания, максимальное число альтернативных путей или список узлов со свободными ресурсами. **Результатом процесса маршрутизации на постоянной основе является четкий маршрут, удовлетворяющий требованиям организации виртуального канала запрашиваемым соответствующим узлом сети.** Четкий путь для каждого маршрута трафика — это список узлов с жестко предустановленными параметрами, который используется для спецификации пути LSP.

С внедрением маршрутизации на постоянной основе ассоциировано с внедрением целого ряда нюансов. Это механизмы обмена информацией:

- Об изменениях топологии сети.
- О доступных ресурсах в узлах сети.
- О состоянии физических линий.
- О распределении ресурсов между процессами маршрутизации на постоянной основе.

Для обеспечения этой информацией, существующий протокол Интернет шлюзования ЮР может быть расширен для включения описания каждого соединения, как пути каждого маршрутизируемого соединения. На практике администратор сетевых ресурсов оператора мультисервисной сети будет описывать конечную точку (и может быть несколько промежуточных узлов) виртуального пути и устанавливать его параметры, которые будут соответствовать требованиям QoS данного информационного потока. **Эта информация направляется в стартовый узел прокладки виртуального пути для маршрутизации информационного потока на постоянной основе, которая должна гарантировать его транспортировку в соответствии с требованиями QoS. Найденный маршрут используется RSVP для прокладки пути LSP.**

В случае перегрузки в сети, таблицы маршрутизации должны быть пере-строены для восстановления всех соединений с помощью автономных систем. Это относительно медленный процесс, сильно зависящий от разрыва соединений и нестабильности маршрутизации как внутри, так и вне узла сети. **Предустановка пути для доставки необходимой информации от MPLS баскир делает возможным быстрое восстановление. В случае сбоя входного или выходного узла пути коммутации меток LSP трафик переадресуется на пустующий, альтернативный путь. При этом нет необходимости ожидать, пока все узлы переопределят свои таблицы маршрутизации для восстановления гарантий QoS соответствующих приложений.** Этот процесс может быть дополнен резервированием на физическом уровне.

2.5.5. Сравнительный анализ механизмов QoS мультисервисных сетей

Обзор механизмов обеспечения QoS в мультисервисных сетях показывает, что для гарантии качества услуг приложений, использующих информационные потоки реального времени, **наиболее важными являются механизмы, реализующие абсолютную схему QoS. Эти механизмы реализуют принципы маршрутизации трафика по сетям виртуальных каналов, которые в сети с дайтаграммной маршрутизацией эмулируют виртуальные пути.**

Обзор абсолютных механизмов QoS, разрабатываемых для IP-сетей регионального масштаба, показывает, что, несмотря на титанические усилия по привнесению эффективных механизмов ATM в среду *IP*, **их дайтаграмная природа существенно затрудняет определение значений параметров виртуальных каналов, гарантирующих качество услуг приложений реального времени. Переменный размер IP-пакетов определяет вероятностный характер оценок параметров качества, что существенно затрудняет исчисление минимально допустимых значений предустановок в узлах сети.**

Таким образом, сеть MPLS способна обеспечить передачу интегрального трафика с «жестким качеством обслуживания» QoS. Технология MPLS хорошо подходит в качестве основы для моделирования трафика, поскольку позволяет сетевым администраторам указывать точный физический маршрут для маркированных пакетов, а также выбирать маршруты, соответствующие специфическим требованиям. Моделирование трафика позволяет поставщику сетевых услуг направлять потоки данных не *по кратчайшему пути, вычисленному с помощью традиционного протокола маршрутизации IGP, а через менее загруженные узлы и каналы связи. В домене MPLS для каждого класса FEC может быть организована виртуальная частная сеть VPN.* В этом случае в домене MPLS формируются таблицы маршрутизации отдельно для маршрутизаторов, к которым подключены только пользователи VPN со своими IP-адресами. Технология MPLS позволяет инкапсулировать в свои сервисные примитивы любые протокольные блоки сетевого уровня. Технология MPLS может оказать помощь провайдерам при внедрении DiffServ. В сетях MPLS с поддержкой DiffServ могут быть организованы два типа LSP: Exp-Label Switch Path (E-LSP) и Label-Label Switch Path (L-LSP). E-LSP - это путь, содержащий несколько различных требований QoS LSP, которые обеспечивают до 8 различных классов SLA. Значение DSCP IP пакетов, для входного узла пути LSP, полностью размещается в поле заголовка MPLS. В поле расширения переносятся как алгоритмы обработки запросов, так и процедуры транспортировки пакетов. Аналогично тому, как в сети с поддержкой DiffServ устанавливается соответствие между алгоритмами PHB и соответствующим классом трафика, определяемым DSCP, в DiffServ домене MPLS устанавливается соответствие между алгоритмом PHB и полем расширения, носящим локальный характер, определяемым сервис-провайдером и конфигурирующим каждый путь LSP. Коммутируемый путь от метки к метке (L-LSP) описывает путь LSP информационного потока с одним набором параметров QoS. Соответствие классов QoS определяется метками MPLS, поэтому возможно определить алгоритм PHB в поле расширения. В этом заключается различие между метками поля расширения и PHB. Пути E-LSPs и L-LSPs могут быть установлены с резервированием полосы пропускания или без него.

MPLS позволяет сервис-провайдеру организовать предоставление услуг VPN, используя простой, гибкий и мощный механизм туннелирования. *Виртуальная частная сеть строится как совокупность маркированных маршрутов между различными физическими сегментами VPN. Система маршрутизаторов провайдера распределяет по всей сети информацию о масках подсетей, существующих внутри каждого сегмента. Входные LSR-маршрутизаторы сети направляют трафик VPN по соответствующим LSP-маршрутам, исходя из совокупности адреса назначения пакета и его принадлежности к определенной VPN.* Хотя технология MPLS VPN описана в рамках общих стандартов на MPLS, в организации IETF существуют специальные рабочие группы, занимающиеся проработкой отдельных вопросов MPLS VPN. Например, рабочая группа PPVPN (Provider Provisioned Virtual Private Network) занята созданием “каркаса” VPN с улучшенными механизмами безопасности, масштабируемости и управляемости (<http://www.ietf.org/html.charters/ppvpn-charter.html>).

Три основных типа VPN, над которыми работает группа PPVPN, — это MPLS BGP VPN, MPLS Virtual Routers и MPLS Layer 2 VPN. Группа координирует свою деятельность с другой рабочей группой — PWE3 (Pseudo Wire Emulation Edge to Edge), создающей стандарты для туннельных сквозных соединений через сети ATM и MPLS на первом и втором уровнях (<http://www.ietf.org/html.charters/pwe3-charter.html>).

И наконец, виртуальные частные сети второго уровня (Layer 2 VPN) также определены в проекте, который получил название Martini, и этот проект тоже сейчас находится на рассмотрении рабочей группы IETF PWE3. Идея здесь заключается в организации туннелей для трафика Ethernet, Frame Relay, ATM и PPP через сеть MPLS. Группа PWE3 работает и над другими похожими предложениями, но со стороны сервис-провайдеров наибольший интерес вызвал проект Martini.

Давайте посмотрим, как работает сеть MPLS BGP VPN, определенная в документе RFC 2547bis. Сразу заметим, что технологии, используемые на первом и втором уровнях, в этом примере нам не важны.

На рисунке показаны *пограничные клиентские маршрутизаторы (CE — Customer Edge)*, которые терминируют протоколы первого, второго и третьего уровней на границе клиентской сети; пограничные маршрутизаторы провайдера (PE — Provider Edge), находящиеся на границе сети сервис-провайдера и выполняющие основную работу, и, наконец, магистральные маршрутизаторы сети провайдера, поддерживающие MPLS LSP.

Маршрутизатор CE использует либо статическую маршрутизацию, либо протокол маршрутизации класса IGP (Interior Gateway Protocol), скажем RIP, OSPF, IS-IS (Intermediate System to Intermediate System), EGP (Exterior Gateway Protocol) либо BGP.

Какой конкретно протокол будет использоваться, в общем-то, безразлично. Для корпоративных пользователей это означает, что не нужно ничего менять в настройках и способах доступа — надо лишь установить соединения на третьем уровне с PE-маршрутизатором. Поставщикам же услуг не придется возиться с сотнями маршрутизаторов пользователей. Именно это мы и имели в виду выше, говоря *о четком определении границ управления в инфраструктуре MPLS.*

Для каждого интерфейса или субинтерфейса (subinterface) на PE-маршрутизаторе формируется своя таблица маршрутизации VRF (VPN Routing and Forwarding). Она является основным элементом MPLS BGP VPN. В ней хранятся данные о маршрутах, информация о которых получена от клиентских маршрутизаторов. В нашем примере SE-маршрутизатор Банка 1 сообщает маршрутизатору PE-1 свой диапазон адресов 10.10.1.0/24, которые заносятся в VRF-Bank. Запись VRF-Bank содержит сведения об изначальном маршруте IPv4 вместе с дескриптором маршрута (Route Descriptor — RD). Поля RD должны содержать номер автономной системы (ASN); именно RD гарантирует, что маршрут уникален. Комбинацию IPv4 и RD называют адресом VPN-IPv4. Обычно создается именно VPN-IPv4 адрес, но в документах RFC 2858 упоминаются еще IPv6 и IPX.

Пограничные маршрутизаторы провайдера распределяют маршруты из VRF с помощью протокола MP-BGP (Multiprotocol BGP). Этот протокол обладает обратной совместимостью с обычным вариантом BGP, однако последний, в свою очередь, конечно, не может распределять адреса VPN-IPv4.

На маршрутизаторе PE-1 в записи VRF-Bank имеется специальный атрибут — Export Target. Он определяет, какой PE-маршрутизатор получит распределяемый с помощью BGP маршрут VPN-IPv4, указанный в VRF-Bank. Аналогично на PE-маршрутизаторах, являющихся получателями этого маршрута, в соответствующих VRF-таблицах должен быть указан объект импорта (Import Target). В нашем примере этим объектом является PE-2 VRF-Bank. Как только информация о маршруте распределена и принята маршрутизаторами, может начинаться обмен данными.

Заметим, что PE-маршрутизаторы содержат данные только о тех SE-маршрутах, которые определены в их VRF-таблицах (информации обо всех маршрутах ко всем

СЕ-маршрутизаторам в сети они не имеют). Точно так же магистральные маршрутизаторы “знают” только о РЕ-маршрутах. Это сильно упрощает работу маршрутизаторов.

Технологией MPLS предусмотрено использование стека меток. Когда пакет покидает СЕ-маршрутизатор Банка 1 (CE Bank 1) и отправляется в путь к пограничному маршрутизатору Банка 2 (CE Bank 2), он сначала поступает на маршрутизатор РЕ-1. “Зная” интерфейс, через который поступил пакет, маршрутизатор РЕ-1 ищет маршрут VRF. Когда найдено соответствие, метка MPLS, относящаяся к следующему в очереди пересылки РЕ-устройству (в данном случае РЕ-2) заносится в стек. Затем РЕ-1 помещает в верхнюю позицию стека метку, определяющую следующую пересылку — магистральный маршрутизатор Р-1. Получив пакет, тот изучает метку, определяет дальнейшую пересылку и помещает метку на следующий узел — маршрутизатор Р-2 (при этом метка РЕ-2 остается в стеке).

Описанный процесс продолжается до последнего магистрального маршрутизатора, в нашем примере это Р-3. Изучая метку РЕ-2, он “понимает”, что маршрутизатор РЕ-2 подключен к нему напрямую и направляет пакет на данный пограничный маршрутизатор. РЕ-2 удаляет метку и перенаправляет изначальный пакет IPv4 на интерфейс пограничного маршрутизатора Банка 2 (CE Bank 2). В обратную сторону трафик передается тем же способом, только используются другие метки. При этом создается и другой LSP-путь, который может проходить через те же магистральные устройства (но это совсем не обязательно). Вы можете создать ячеистую сеть или сеть типа “звезда”, изменяя объекты импорта и экспорта.

Обратите внимание на то, что пакеты в нашем примере проходят через сеть провайдера при помощи “промежуточных” (shimmed) меток. IP-адресация не используется, поэтому механизмы трансляции сетевых адресов (NAT) не нужны. Это означает, что сети с дублирующимися адресами (в соответствии с RFC 1918), вроде 10.10.0.0 и 192.168.0.0, будут нормально взаимодействовать через инфраструктуру сервис-провайдера.

Вопросы для самопроверки по теме 3.

1. Приведите определение NGN-сети. Укажите основные принципы функциональной архитектуры NGN
2. Сформулируйте основные отличительные черты базовой эталонной модели NGN
3. Дайте общую характеристику и особенности инфоуслуг
4. Приведите примеры наиболее популярных предоставляемых и потенциальных услуг технологии Triple Play
5. Опишите основные характеристики базовых потоков информации в ИКС
6. Перечислите основные компоненты и их назначение мультисервисной сети связи как области взаимодействия ИКС
7. Укажите службы и специфику функционально-структурной организации инфотелекоммуникационных транспортных IP-QoS-систем. Основные недостатки технологий
8. Укажите службы и специфику функционально-структурной организации инфотелекоммуникационных транспортных ATM-систем
9. Приведите классификацию инфотелекоммуникационных транспортных систем МСС по способу закрепления физических ресурсов сети за соединением
10. Укажите особенности передачи речевых сигналов в пакетных ИТС. Обеспечение QoS-норм в пакетной и гибридной транспортной системе в режиме сессии

11. Охарактеризуйте МСС как область взаимодействия инфокоммуникационной сети как интеллектуальной компьютерной системы. Сформулируйте определение инфотелекоммуникационной транспортной системы.
12. Приведите классификацию ИТС по способу закрепления физических ресурсов сети за мультимедийным соединением.
13. Назовите основные требования, предъявляемые к передаче различных типов информации в инфокоммуникационной транспортной системе. Укажите особенности передачи мультимедийных информационных потоков.
14. Назовите основные составляющие общей сетевой задержки трафика классов *B* и *C*. Охарактеризуйте переменную и постоянную составляющие сетевой задержки.
15. Сформулируйте определение изохронности потока.
16. Какие способы обеспечения изохронности передачи трафика Вы знаете?
17. Сформулируйте основные принципы построения МСС.
18. Обоснуйте показатели качества в МСС.
19. Сформулируйте общие требования, предъявляемые к критерию эффективности сложных систем.
20. Сформулируйте общие подходы к выбору критерия эффективности МСС.
21. Обоснуйте функциональный критерий эффективности ИТС.
22. Сформулируйте задачу структурного синтеза при оптимизации по максимуму критерия эффективности.
23. Приведите в общем виде выражения проектных процедур параметрического анализа, параметрического синтеза и параметрической оптимизации в терминах критериев эффективности ИТС.
24. Опишите концепцию архитектуры МСС и ее основные особенности
- 25.

ТЕМА 4. МЕТОДОЛОГИЯ МОДЕЛИРОВАНИЯ И АНАЛИЗА ПАКЕТНОЙ ИНФОТЕЛЕКОММУНИКАЦИОННОЙ ТРАНСПОРТНОЙ СИСТЕМЫ

3.1. Обоснование качественных показателей обслуживания мультимедийного трафика МСС

Состав показателей качества обслуживания QoS, ограничения на значения которых должны учитываться в процессе создания ИТС, подробно рассмотрены в [8.23, 8.24]. Рассмотрим конкретный состав показателей качества обслуживания мультимедийных потоков в режиме установленного соединения и обоснуем набор ограничений на них, обусловливаемых требованиями пользователей и/или нормативными документами.

При создании ИТС должны, в первую очередь, выполняться ограничения на значения показателей качества обслуживания вызовов различных типов пользователей на фазе установления соединения, а именно, вероятность потери мультимедийного вызова

$$b^{multy} = b(y^{multy}), \quad (4.1)$$

нормативные (допустимые) значения которых определены соответствующими рекомендациями ИТУ. Здесь y^{multy} – интенсивность поступления мультимедийных вызовов. Создание методики расчета потерь в ИКС – задача теории телетрафика будущего. Отметим, что этот параметр в существующих моделях сетей указанного класса или ее элементов, как правило, не учитывается. Как отмечено выше, в роли ограничений на значения показателей качества обслуживания изохронного трафика в сессии в существующих моделях сетей с интеграцией служб выступают либо средняя величина сетевой задержки пакета изохронного трафика T^B , либо вероятность $P_r\{t \leq \theta_{st}^B\}$ доставки пакета указанного класса в сети за время, не превышающее заданное θ_{st}^B . В дальнейших исследованиях мы будем применять «квантильный» подход, т. е. будем вводить ограничение на заданную вероятность d^B этого превышения, которая гарантирует необходимую изохронность передачи [8.23, 8.24]

$$P_r\{t \geq \theta_{st}^B\} \leq d^B \quad (4.2)$$

Это связано со спецификой обслуживания любых изохронных информационных потоков, о чем речь пойдет ниже. С целью уточнения понятия блокировки соединений в сессии введем понятие блокировки ресурса на примере СМО общего типа $G/G/m/w$ [8.55]. Введем событие: « σ^k заявка класса k принята на обслуживание и время ожидания ее обслуживания не превосходит t^k, c ». Указанное событие означает, что для обслуживания заявки ресурса СМО достаточно. Обозначим это событие $\{\sigma^k; t^k\}$. Пусть вероятность этого события есть $\Pr\{\sigma^k; t^k\}$. Вероятность дополнительного события – «заявка σ^k не принята на обслуживание или если она принята, то время ожидания ее

обслуживания больше t^k, c », - есть $1 - \Pr\{\sigma^k, t^k\}$. Назовем эту вероятность вероятностью нехватки или блокировки ресурса P_{bl} СМО и обозначим $P_{bl} \equiv p(m, w, t) = 1 - \Pr\{\sigma^k, t^k\} = 1 - \Pr\{\sigma^k\} \Pr\{t^k | \sigma^k\}$, где $\Pr\{\sigma^k\}$ - безусловная вероятность того, что заявка принята на обслуживание; $\Pr\{t^k | \sigma^k\}$ - вероятность того, что время ожидания заявки класса k не превосходит t^k, c при условии, что она принята на обслуживание. Таким образом, вероятность $p(m, w, x)$ для СМО типа $G/M/m/w$ определяет вероятность блокировки всего ресурса. Если обозначить через p_{m+w} - вероятность того, что в системе все приборы и места для ожидания заняты (произвольный вызов теряется), $G^*(t)$ - условная функция распределения времени пребывания заявки в системе при условии, что она не теряется, то $P_{bl} \equiv p(m, w, t) = 1 - (1 - p_{m+w})G^*(t)$. Точные аналитические выражения p_{m+w} и $G^*(t)$ для СМО типа $G/M/m/w$ приведены в [8.56]. В частности, для этой системы $p(m, 0, 0)$ - вероятность потери заявки в СМО типа $G/M/m/0$, а $p(1, \infty, \theta_{st}^B)$ - вероятность превышения времени пребывания заявки θ^B в СМО типа $G/M/1/\infty$.

Действительно, как следует из [99], при $w = 0 - G^*(t) = \delta(t)$ (функция единичного скачка) и мы приходим к обычному определению потерь в m -линейных системах СМО без памяти. Если $w = \infty$, то $p_{m+w} = 0$ и вероятность блокировки ресурса системы дается выражением $P_{bl} \equiv 1 - G^*(\theta^B)$.

Последний случай соответствует моделям СМО, традиционно применяемым для анализа сетей с коммутацией пакетов. Таким образом, для пакетных ТС, отдельные звенья которых моделируются СМО типа $G/M/1/\infty$, понятие вероятности блокировки ресурса в сеансе связи для речевого трафика класса B становится эквивалентным вероятности превышения B -пакетами заданного сквозного времени пребывания в сети θ^B . Обозначим эту вероятность d^B . Эта величина характеризует качество переноса речевых B -пакетов. В физическом смысле вероятность d^B есть доля пакетов класса B , превысивших время θ^B и/или потерянных из-за ошибок в заголовке и переполнения буферов маршрутизаторов [8.26]. Другими словами, для качественного воспроизведения речи важно не среднее время пребывания речевого пакета в сети, а доля речевых пакетов, не доставленных получателю за заданное время, т. е. при анализе необходима фиксация заданного квантиля распределения времени их пребывания в тракте передачи.

Качественный показатель для оценки межпоточной синхронизации передачи изохронного потока с более строгими требованиями к изохронности передачи (например, видеопоток) по сравнению с потоком с менее строгими требованиями (например, аудиопоток) будем оценивать коэффициентом межпоточного смещения

$$R_{skew} = \frac{d^{k1}}{d^{k2}} \leq const, \quad (4.3)$$

где d^{k1} и d^{k2} - заданный уровень изохронности передачи потоков k -го класса.

Для характеристики качества передачи асинхронного трафика, вводится заданное среднее время пребывания пакета данных в трактах сети. Для характеристики качества передачи асинхронного трафика в ИТС, вводится заданное среднее время пребывания

пакета данных $T_{st,m}^{Cp} = \int_0^{\infty} t \varphi_{st,m}^{Cp}(t) dt \leq T^{Cp}$ в m -м маршруте тракта $st \in S^{Cp}$, не превосходящее предельной величины T^{Cp} , где $\varphi_{st,m}^{Cp}(t)$ – плотность вероятности времени пребывания пакета данных в m -м маршруте тракта $st \in S^{Cp}$, т. е.

$$T_{st,m}^{Cp} \leq T^{Cp} \quad (4.4)$$

Это связано с тем, что (в отличие от обычно применяемой в существующих моделях среднесетевой задержки) для пользователей сети представляет интерес не просто минимальное время пребывания пакета в сети (которое само по себе может оказаться достаточно большим и неприемлемым, например, для интерактивного обмена), а заданное среднее время. При более жестком нормировании качества переноса высокоскоростных потоков данных класса C (см. табл.), связанным с ограничением на долю потерянных и/или засланных не по адресу пакетов данных будем также применять «квантильный» подход, т. е.

$$P_r \{ t \geq T^{C3} \} \leq d^C \quad (4.5)$$

3.2. Обоснование критерия эффективности

В качестве основополагающего методологического средства оценки системы принята ее экономическая и/или функциональная «эффективность». Оценка эффективности сложных систем, к которым относятся ИТС, является сложной и не до конца решенной проблемой. Наиболее важным требованием к критерию эффективности является *способность критерия измерять эффективность рассматриваемой системы*. В этой связи критерий эффективности должен отражать основное назначение системы, исходя из цели ее проектирования. Другим, не менее важным, требованием к критерию является *возможность его оценки в количественной, численной форме*, исключающей неоднозначность толкования его значения. При этом критерий эффективности должен иметь четкий физический смысл и давать наглядное представление об исследуемом качестве системы. Третье по важности требование заключается в *чувствительности критерия* к основным варьируемым параметрам системы, что позволяет определить наиболее рациональные характеристики системы или ее составляющих. Четвертое требование - критерий *должен обеспечить сравнение различных вариантов построения системы* для выбора наиболее рационального. Последнее требование относится к полноте учитываемых критерием главных параметров системы или ее составляющих, определяющих назначение и качество работы системы. Кроме того, критерий *должен быть достаточно простым и удобным в инженерной практике расчета*. Эффективность системы, а, следовательно, и ее критерий в общем случае зависят от трех групп факторов: характеристик многомерного входящего потока (интенсивности, длин сообщений и др.); особенностей функционирования системы; экономических показателей, характеризующих ее работу и др.

Как отмечалось выше, основным характеристическим показателем качества функционирования (характеристической мерой) ИТС с учетом ее назначения является степень использования пропускной способности системы, а конечной **целью** функционирования - максимально эффективное использование ее ресурсов при организации мультимедийного соединения. Это, в свою очередь, порождает две целеполагающие функции системы. Первая функция связана непосредственно с самой системой и характеризует эффективность или степень использования ресурса системы, т. е. является характеристикой качества функционирования системы (Network

Performance, NP). Вторая функция связана с процессом переноса мультимедийной информации и характеризует качество обслуживания пользователя QoS, т. е. является характеристикой качества обслуживания.

Назначение ИТС **однозначно предопределяет** вид ее комплексного функционального критерия эффективности системы (целевой функции) K в виде общей числовой характеристики использования пропускной способности системы мультимедийным трафиком с учетом ограничений на QoS-нормы его переноса. Указанная скорость определяется номинальными скоростями отдельных потоков, обслуженных системой и дается аддитивной формой

$$V_{st}^{\min} = \sum_k V_{st}^{k \min} = \sum_k K_{st}^k V_{st} \quad (4.6)$$

$$\text{при } P_r \{ t \geq \theta_{st}^B \} \leq d^B, R_{skew} = \frac{d^{B1}}{d^{B2}} \leq const, T_{st}^C \leq T_{\max}^C (P_r \{ t \geq T_{st}^C \} \leq d^C) \quad (4.7)$$

При этом значение номинальной полосы пропускания должно удовлетворять ограничениям допуска нового соединения, в качестве которых могут выступать ограничения на качественные показатели транспортировки разнородного трафика в сессии. Оценку эффективности функционирования пакетной ИТС в режиме установленного соединения можно оценить также коэффициентом использования сквозного цифрового тракта передачи трафиком различной природы на транспортном уровне:

$$R^P = [V_{st}^{\min B} (1 - d^B) + V_{st}^{\min C} (1 - d^C)] / V_{st} \dots \dots \quad (4.8)$$

Та система лучше, у которой этот коэффициент лучше при заданных условиях проектирования. Расчет показателей пропускной способности и качества обслуживания сети связи относится к задачам анализа.

Критерий K как характеристическая мера является первичным критерием эффективности, в то время как вторичные критерии порождают ограничения на качественные показатели переноса мультимедиа. Необходимо отметить, что построение удобного в инженерной практике комплексного показателя эффективности, объединяющего в себе основные частные критерии является не тривиальной задачей.

3.3. *Концептуальная модель инфотелекоммуникационной транспортной системы*

Под методологией анализа понимают совокупность системных методов, указывающих возможные альтернативные способы достижения цели и принципов, направленных на решение проблемы моделирования и исследование процессов функционирования ИТС. При этом принцип, как один из аспектов методологии, — есть постоянно и последовательно применяемый метод.

При моделировании МСС и, в частности, ее ИТС, важнейшей задачей остается разработка комплексной математической модели системы, позволяющей связать экономические и структурные ее характеристики с характеристиками процесса доставки сообщений, надежности, управления и технического обслуживания. На основе этой модели разрабатываются методы проектирования МСС с использованием средств автоматизированного проектирования (САПР) и системного подхода ко всему циклу разработки сети.

Описание ИТС и условий ее функционирования характеризуется определенной совокупностью параметров. При анализе и синтезе любой системы формируются, в основном, три группы данных: 1) входные параметры, определяющие ограничения задачи оптимизации; 2) внутренние или проектные параметры, под которыми понимают

независимые переменные параметры, полностью и однозначно определяющие решаемую задачу проектирования; 3) выходные или целевые параметры (интегральные и дифференциальные): экономические, вероятностно-временные характеристики процесса доставки сообщений, использования сетевых ресурсов и др. Внешние и внутренние группы параметров образуют параметрические базисы [85, 188].

Среди выходов могут быть выделены: переменные — критерии, максимизируемые или минимизируемые в процессе оптимизации, переменные — лимиттеры, на которые накладываются ограничения. Внутренние переменные применительно к задаче оптимизации разбиваются на две группы — управляемые и неуправляемые. Первые — модельные параметры, непосредственно влияя на которые алгоритм осуществляет оптимизацию, вторые — различные производные от управляемых, которые могут быть как контролируемыми, так и неконтролируемыми. На переменные этого базиса также могут накладываться ограничения.

Входные параметры ИТС определяют ограничения задач оптимизации и задаются вектором I^k , который может быть представлен набором следующих агрегатов:

$I^k = [Q^k, \lambda_{ij}^k, Y^k, \rho_{ij}^k, v^B, \omega^C, s^C, \tau^B]$. Здесь Q^k количество конечных систем (ES) в k -ом классе трафика. Нагрузочные характеристики: λ_{ij}^k — суммарная интенсивность входного потока k -го класса в линейно-цифровых трактах (ЛЦТ) инцидентных узлов коммутации $-ij$, пакет/с (сообщений/с); Y^k, ρ_{ij}^k — соответственно суммарная нагрузка, поступающая в систему, (эрл) и коэффициенты межузловой загрузки ЛЦТ; v^B — скорость работы речепреобразующего устройства, бит/с; ω^C — скорость работы абонентской установки данных, бит/с; s^C — длины сообщений данных, бит; τ^B — длительность речевого фрагмента, с; $c_{УК}$ — производительность УК, пакет/с.

Внутренние параметры ИТС могут быть представлены следующим вектором: $W^k = [G, H_h^k, S_h^k, A^{*k}, d_s^{A(B)}, T_{st}^C, V_{ij}, V^{k,u}, p_{ij}^{ou}, L^k]$, где подвектор G отображает тип топологической структуры; подвектор протоколов $H_h^k = [H_{Ph}^k, H_L^k, H_N^k, H_T^k]$ отображает типы протоколов ИТС ($h_p = \overline{1,4}$ — для пакетных ИТС на технологии IP-QoS, $h_p = \overline{1,5}$ — для ИТС на технологии ATM); подвектор $A^{*k} = [A^k, S_h^k(M_h)]$ отображает тип административной системы управления, в том числе системы управления безопасностью, где A^k — характеризует систему сетевого управления, а $S_h^k(M_h)$ — базовые S -услуги безопасности, реализуемые M — механизмами защиты на h -уровне логической структуры сети ($S = \overline{1,5}$, $M = \overline{1,14}$ [12]); p_{ij}^{ou} — вероятность ошибки в ЛЦТ $ij \in I^k$; V_{ij} — скорость передачи в ЛЦТ, бит/с; L_h^k — длина протокольного h -уровневого примитива логической структуры сети, бит;

Выходные или целевые параметры ИТС задаются вектором $O^k = [C^k, P^k, K^k]$. Подвектор $P^k [b^k, d^B, R_{sket}, T^C]$ отображает вероятностно-временные характеристики ИТС, определение компонент которого дано выше; подвектор $C^k [C_K^k, C_{II}^k, D_{Пог}^k]$ отображает стоимостные характеристики и $K^k [I^k, W^k]$ функциональные характеристики. При необходимости задаются также и требования к этим характеристикам. Среди экономических критериев предпочтение отдается приведенным затратам [130 – 132].

Абстрактную модель изучаемой системы в самом общем виде можно представить в виде зависимости

$$O^k = f(I^k, W^k), \quad (4.9)$$

где O^k — некоторый выходной (целевой) количественный показатель эффективности системы в плане достижения цели ее существования V^{\min} , будем называть его — критерий эффективности (в нашем обозначении K^k); W^k — управляемые переменные системы; I^k — неуправляемые внешние по отношению к системе воздействия.

Указанная функциональная зависимость является в общем виде концептуальной или системной моделью ИТС или ее критерием эффективности в плане достижения ее целеполагания. Подобная формализация необходима для декомпозиции общей задачи анализа процессов функционирования ИТС на частные задачи и перехода от внешнего описания системы к описанию ее внутреннего строения. При этом возможны следующие задачи [55, 85, 110, 189]:

$$\text{Дано: } I^k, W^k. \text{ Найти } O^k \quad (4.10)$$

$$\text{Дано: } I^k, \text{ ограничения на } O^k. \text{ Найти } W^k. \quad (4.11)$$

$$\text{Дано: } I^k, \text{ ограничения на } O_1^k. \text{ Найти } W_0^k, \text{ доставляющее } \textit{opt} O_2^k; O^k = [O_1^k, O_2^k]. \quad (4.12)$$

Если фиксировать компоненты вектора W^k и варьировать их параметрами, то получим соответственно следующие проектные процедуры: параметрического анализа (4.10) — (заданы структура сети и пропускные способности пучков каналов; требуется минимизировать ВВХ сети, варьируя потоками), параметрического синтеза (4.11) — (заданы структура сети, матрица нагрузок, максимальная общая стоимость сети; требуется минимизировать ВВХ сети, варьируя пропускной способностью отдельных пучков каналов) и параметрической оптимизации (4.12). В задаче (4.12) вектор O_2^k — количественный критерий эффективности системы, а компонентами O_1^k являются ограничения оптимизационных задач на вероятностно-временные характеристики ИТС. Выбор состава оптимизируемых параметров и разработка математических методов расчета показателей качества и критерия эффективности защищенной ИТС основаны на ряде предположений о функционально-структурной организации подсистем. Для оценки эффективности и расчета вероятностно-временных характеристик (ВВХ) системы необходимо: 1) построить критерий эффективности ИТС; 2) разработать на базе критерия эффективности ее аналитическую модель, выявляющую функциональные зависимости критерия эффективности и критерия обслуживания от параметров системы; 3) разработать алгоритмы численного расчета характеристик системы на основе полученных аналитических зависимостей.

Вопросы для самопроверки по теме 4.

ТЕМА 5. ПОСТРОЕНИЕ МОДЕЛЕЙ ПАКЕТНОЙ ИНФОТЕЛЕКОММУНИКАЦИОННОЙ ТРАНСПОРТНОЙ СИСТЕМЫ

5.1. Концепция архитектуры мультисервисной сети и ее основные особенности

Под архитектурой МСС будем понимать уровневую организацию необходимых функций, протоколов и их программных реализаций, которые должны выполняться сетью для целенаправленного *инфокоммуникационного обслуживания мультимедийных компонентов прикладного уровня* (предоставления инфоуслуг). Построение и исследование архитектуры МСС позволяет:

- иерархически разделить сетевые функции, на логически завершенные функциональные группы задач (логические уровни или подсистемы), для определения механизмов их реализации;
- выявить особенности их функционирования, связанные с обслуживанием мультимедийного трафика в рамках единой транспортной услуги;
- выработать единый язык для описания множества функций взаимодействия в сети;
- детализировать протоколы и форматы протокольных блоков соответствующих логических уровней применительно к передаче и обработке мультимедийного трафика;
- провести комплексный анализ и синтез систем на основе декомпозиции их функциональных и протокольных структур.
- разработать модели, методы и алгоритмы анализа и синтеза различных реализаций инфотелекоммуникационных систем;
- объединить задачи по техническому обслуживанию и управлению сети и ее элементов и др.

Построение моделей процессов функционирования МСС требует наличия специальной терминологии для формализованного описания объектов исследования. Для формализованного описания функционально-структурной организации и выявления общих принципов построения ИТС будем использовать терминологию и принципы формализованного описания объектов исследования модели архитектуры взаимодействия открытых систем (ВОС) - (Open System Interconnected, OSI), сформулированные в рекомендациях серии I. МККТТ и в эталонной модели (ЭМ) Международной организации стандартов (МОС) - (International Standard Organization, ISO) и получившие дальнейшее

развитие для моделей архитектур ЦСИС. Специфику МСС при ее проектировании важно выявить в терминах этих моделей, учитывающих специфику передаваемой информации, использования ресурсов сети и протоколов обмена различными видами информации. При этом мы будем рассматривать МСС как более широкий класс открытых сетей.

Архитектура МСС описывает построение и взаимосвязь ее логической, программной и физической структур.

Логическая структура сети описывает ее полный функциональный профиль и базируется на рассредоточении процессов передачи и обработки разнородной информации по функциональным (логическим) h -уровням каждой из ее систем и имеет многослойный вид. Система в модели OSI - это автономное объединение одной или нескольких ЭВМ, соответствующего программного обеспечения, периферийных устройств и пользователей, способное обрабатывать и передавать информацию. Систему, удовлетворяющую требованиям стандартов ISO, именуют открытой. Отметим, что модель OSI описывает взаимодействие открытых систем и не описывает внутренние функции, выполняемые конкретной системой. Логическая структура сети должна удовлетворять принципам системной декомпозиции, в частности, обеспечивать относительную независимость уровней друг от друга, что позволяет модифицировать функции любого уровня, определить тип структурных отношений, характеризуемых упорядоченностью и организованностью взаимодействий между отдельными уровнями по вертикали, а также разложить ИТС по парам «источник-получатель».

Каждая система в МСС рассматривается в виде логически взаимосвязанной совокупности подсистем, образованных в результате пересечения системы с некоторым h -уровнем. Уровень образуют подсистемы одного ранга. Каждая подсистема состоит из одного или нескольких h -уровневых объектов. Суть уровневой организации состоит в предоставлении более высокому смежному логическому уровню сервисных услуг более низкого логического уровня. Услуга - это функциональные возможности h -уровня, которые предоставляются в распоряжение ($h+1$) - объектам в h -точках доступа услуг (h -ТДУ), которые играют роль логических интерфейсов (правил взаимодействия между смежными уровнями) между h -объектами и $h+1$ -объектами. При создании однородных ИТКС, например, на базе волоконно-оптических линий связи, методика логического проектирования значительно упрощается за счет упразднения отдельных функциональных уровней архитектуры сети.

Отличительной особенностью архитектуры сетей указанного класса является наличие в их логической структуре, по крайней мере, трех базовых дополнительных функций: функции «управления резервированием сетевых ресурсов», функции «контроля допустимости установления сессии» и функции «совмещения», а также функций безопасности с учетом классов трафика.

Программная структура МСС реализует функциональный профиль сети, представленный объектами логических уровней, базируется на декомпозиционной иерархии ее программного обеспечения и описывает взаимодействие связанных между собой отдельных программ, отображающих работу и взаимосвязь логических уровней. Одноранговые объекты h -уровня взаимодействуют между собой с помощью одного или нескольких протоколов через логические соединения, создаваемые на $h-1$ уровне. Спецификация протоколов h -уровня определяет процедуры выполнения служб, форматы управляющих и информационных полей протокольных блоков уровня, процедуры обмена протокольными блоками между объектами h -уровня в разных открытых системах, а также механизм выбора указанных процедур из списка возможных. Протокольным блоком данных уровня (protocol data unit, PDU) называются фрагменты информации, пересылаемые между одноранговыми объектами уровня двух систем. Каждый логический h -уровень формирует протокольный блок данных из сервисного блока данных (service

data unit, SDU), переданного вниз с $h+1$ -уровня, добавляя к нему управляющую информацию конкретного протокола взаимодействия одноранговых объектов своего h -уровня (protocol control information, PCI). Часть информации, составляющей PCI, передается с $h+1$ -уровня на h -уровень в виде параметров запроса службы. h -объекты h -уровня взаимодействуют между собой через логические каналы. Например, логические каналы между объектами канального уровня модели OSI именуются каналами передачи данных, а между объектами транспортного уровня – виртуальными каналами.

Физическая структура МСС базируется на конкретных технических устройствах и позволяет оптимально реализовать в них отдельные логические функции или их совокупность.

5.2. *Архитектура инфотелекоммуникационных транспортных систем на технологии IP-QoS*

Модель архитектуры ИТКС на технологии IP-QoS базируется на четырехуровневой концептуальной модели архитектуры сети Агентства перспективных исследований национального Министерства обороны США (Defence Advances Research Agency, DARPA). Логическая и программная структура ИТКС на технологии IP-QoS включает в себя следующие функциональные уровни: прикладной, транспортный, межсетевой и сетевого интерфейса или сетевого доступа. Логическая и программная структура архитектуры ИТКС на технологиях IntServ и DiffServ приведена на рис.5.1. Анализ особенностей моделей архитектуры ИТКС на технологии IP-QoS с позиций размещения, организации и реализации базовых функций и их реализации проведен в [16, 26,27, 78, 79].

Программная структура архитектуры DiffServ	Программная структура архитектуры IntServ	Логическая структура архитектуры МСС IP-QoS	Логическая структура архитектуры DARPA
FTP, Telnet, SMTP DNS, RIP, SNMP, <i>SLA/SLS (драйвер RSVP+), RTP, RTCP, RTSP</i>	FTP, Telnet, SMTP DNS, RIP, SNMP, <i>драйвер RSVP, RTP, RTCP, RTSP</i>	Прикладной А-уровень (<i>функция установления сквозных соединений с заданными QoS-нормами передачи трафика, функция синхронизации мультимедийных потоков</i>)	Прикладной А-уровень
TCP,UDP, <i>RSVP+</i>	TCP,UDP, <i>RSVP</i>	Транспортный Т-уровень (<i>функция управления резервированием сетевых ресурсов, функция «совмещения» трафика, функция синхронизации мультимедийных потоков</i>)	Транспортный Т-уровень (<i>функции транспортного уровня модели ВОС и часть функций ее сеансового уровня</i>)

IP (ARP, IGMP, ICMP), <i>TCA/TCS; DSCP, TC, PHB</i>	IP (ARP, IGMP, ICMP), <i>алгоритмы контроля доступа и политики; ToS, классификатор пакетов, планировщик пакетов</i>	Межсетевой I-уровень (<i>функция контроля допустимости установления соединения, функция «совмещения» трафика</i>)	Межсетевой I-уровень (<i>поддержка адресации IP-пакетов и их маршрутизация</i>)
Ethernet, Token Ring, FDDI, Frame Relay, <i>MPLS, ATM</i>	Ethernet, Token Ring, FDDI, Frame Relay, <i>MPLS, ATM</i>	Сетевого интерфейса NA-уровень (<i>функция поддержания заданных QoS-норм передачи разнородного трафика</i>)	Сетевого интерфейса NA-уровень (<i>передача IP-пакетов в сетевую среду и прием этих пакетов из сетевой среды</i>)
SDH/SONET, xDLC	SDH/SONET, xDLC	Физический Ph-уровень	Физический Ph-уровень

Рис. 5.1 - Логическая и программная структура архитектуры ИТКС IP-QoS

Три нижележащих уровня модели архитектуры DARPA (Т-уровень, I-уровень и NA-уровень) образуют логическую структуру *транспортной системы* сети, которая описывает основные функции собственно сети связи, их логическую взаимосвязь и параметры, поэтому при построении моделей сетей связи целесообразно рассматривать именно свойства и функции их транспортных систем.

Основной задачей транспортного уровня (**Т-уровень**) является обеспечение переноса мультимедийной информации между прикладными объектами оптимальным образом при требуемом качестве обслуживания разнородного трафика. Как правило, функции **Т-уровня** реализуются в ядре операционной системы ПЭВМ пользователя, т. е. **Т-уровень** совпадает с традиционной границей между сферами операторов связи и пользователей. Для организации переноса мультимедийного потока с комплексным представлением информации нескольких типов логические объекты уровня должны обеспечивать установление требуемого количества сетевых логических соединений и задействовать протокол их совместного использования в рамках единой транспортной услуги. Однако может быть режим (более характерный для передачи мультимедиа), когда одно сетевое соединение обеспечивает более чем одно транспортное соединение. В этом случае требуется функция (и протокол) мультиплексирования I-соединения. Очевидно, что транспортный уровень при этом должен иметь информацию о качестве обслуживания каждого I-соединения, и наличии ресурсов I-соединений в каждом направлении. На Т-уровне ИКС на технологии IP-QoS реализуются следующие основные функции:

- создание мультимедийного логического транспортного соединения;
- управления резервированием сетевых ресурсов;
- реализация функции совмещения разнородного трафика;
- обеспечение изохронности переноса элементов потока;
- межпоточная синхронизация мультимедийных потоков;
- управление разнородными потоками информации между двумя портами;

- обеспечение повторной передачи потерянных или ошибочных данных и др.

В сети IntServ функция управления резервированием сетевых ресурсов обеспечивается сигнальным протоколом RSVP. Параметры для резервирования ресурсов содержатся в описателе потока сообщения *Resv*, который включают в себя спецификацию фильтра *filter spec* и спецификацию потока *flowspec*. Спецификацию потока можно разделить на две группы: спецификацию трафика (*Traffic Specification, Tspec*) и спецификацию сервисного запроса (*Service Request Specification, Rspec*). Спецификация *Tspec* определяет количественные параметры потока: среднюю скорость потока; длину пачки пакетов, генерируемой источником; пиковую скорость; минимальный и максимальный размер поля данных IP-пакета PDU, а *Rspec* — параметры качества обслуживания: полосу пропускания, задержку пакета, вероятность потери пакета). Форматы и содержимое «*Tspec*» и «*Rspec*» определяются общими моделями обслуживания [RFC 2210].

Повреждение или фальсификация запросов резервирования может привести к получению услуг неавторизованными пользователями или к отказам в услугах. RSVP осуществляет защиту против таких атак с помощью механизма аутентификации, действующего в каждом из узлов и использующего шифрование с применением хэш-функций. Механизм поддерживается объектами «INTEGRITY», которые могут быть включены в любое сообщение RSVP. Эти объекты используют технику криптографических дайджестов.

В сети DiffServ **функция управления резервированием сетевых ресурсов** реализуется в процессе заключения с пользователем «трафик-контракта» или соглашения на поставку услуги определенного уровня SLA. При этом SLA отведен наивысший уровень абстракции в специфицировании услуги, а параметризация трафика выделена в его самостоятельное подмножество. Параметры потока декларируются в соглашении/спецификации по трафику TCA/TCS (*Traffic Conditioning Agreement*)/(*Traffic Conditioning Specification*). В соответствии с заявленным классом обслуживания приложению осуществляется требуемое резервирование ресурсов. В результате заключения «трафик-контракта» для приложения гарантируется требуемое резервирование сетевых ресурсов и параметризация класса сервиса в сети.

Функция «совмещения» на T-уровне рассматриваемых сетевых архитектур вводится *в явном виде* и реализуется различными программными механизмами в ядре операционной системы ES пользователя. Во-первых, она реализуется здесь «окраской» потока определенными значениями поля «порт получателя» в заголовке TCP-сообщения (Приложение 2) транспортного протокола с контролем передачи (*Transport Control Protocol, TCP*). Во-вторых, для «окраски» потока может быть использовано поле указателя срочности данных UP (*Urgent Pointer*). Это поле содержит номер пакета, начиная с которого следуют пакеты повышенной срочности при установлении флага «срочности» URG в поле «Резерв». Отметим, что каждое соединение однозначно специфицируется парой сокетов (*sockets*), каждый из которых определяется протоколом и адресом хоста. Формат адреса специфичен для каждого протокола. Так, в TCP/IP этот адрес представляет собой комбинацию IP-адреса и порта. Два сокета (по одному на каждой стороне соединения) образуют двухсторонний коммуникационный путь. Механизм сокетов, позволяет работать на одном порту одновременно нескольким приложениям и однозначно идентифицировать каждый поток данных в сети. Существует два типа сокетов: сокет потока (*stream socket*) и сокет дейтаграммы (*datagram socket*).

Основными функциями T-уровня является также поддержание *изохронности* в пределах определенного допуска при передаче изохронного трафика и *межпоточная синхронизация* мультимедийных потоков на входе и выходе транспортного соединения.

Компенсация джиттера на T-уровне может быть осуществлена введением некоторой константы времени задержки в сглаживающем буфере на приеме, либо применением специальных алгоритмов, обеспечивающих выбор и реализацию

оптимальной задержки речевых пакетов в сглаживающем буфере на основе комплексной информации о его прохождении в сети и «синхронизации» часов приемника и передатчика. Для поддержания заданной величины изохронности передачи трафика в конечных устройствах может быть задействован алгоритм виртуального планирования скорости поступления пакетов (ячеек) - (VSA, Virtual Scheduling Algorithm), который является разновидностью общего алгоритма регулирования скорости ячеек (Generic Cell Rate Algorithm – GCRA, определенного в Рекомендации I.371 ITU-T. Для этих целей в конечных системах ES сетей IP-QoS также могут быть задействованы дополнительные механизмы для контроля параметров трафика, такие как, например, протокол реального времени (Real-Time Protocol, RTP) и протокол управления (RTP Control Protocol, RTCP) [RFC1889], используемый совместно с протоколом RTP, а также протокол для потоковых данных реального времени (Real-Time Streaming Protocol, RTSP) [RFC2326]. Семейство рекомендаций ITU-T H.32x описывает функционирование систем мультимедиа связи в различных транспортных технологиях, а семейство рекомендаций T.12x дополняет их в части обеспечения совместной работы над документами.

Межсетевой **I-уровень** обеспечивает установление, поддержание и разъединение логического мультимедийного соединения между двумя объектами транспортного уровня с фиксацией маршрутов переноса многокомпонентных информационных потоков в сессии в рамках единой транспортной услуги. К основным функциям уровня следует отнести:

- контроль допустимости установления соединения;
- совмещение мультимедийного трафика;
- организация маршрутизации и ретрансляции пакетов различных классов трафика в рамках предоставления единой транспортной услуги и др.;

В технологии **IntServ** на I-уровне реализован динамический подход **контроля допустимости установления соединения** САС. Функция САС реализуется в маршрутизаторах посредством двух модулей: модуля «контроль доступа», который на основании параметров flowspec осуществляет проверку наличия требуемых свободных ресурсов и административного модуля «управление политикой», который проверяет полномочия приложения на заявленный ресурс на основании значения данных «POLICY_DATA» из спецификации filter spec. Управление политикой будет зависеть от положительного результата аутентификации для каждого из запросов резервирования. Информация, характеризующая политику, может быть включена в сообщение в виде криптографически защищенного сертификата пользователя. Для параметризации потока используются спецификация фильтра (используется классификатором пакетов для идентификации пакетов одного и того же потока с определенным IP-адресом отправителя и выходным портом) и спецификация потока (используется для задания параметров планировщика пакетов).

Функция контроля допустимости установления соединений в сети IP на технологии **DiffServ** [RFC 2475] фактически реализуется путем соответствующих настроек параметров в базе данных (БД) граничных маршрутизаторов. В любом случае на каждом маршрутизаторе сети классы сервиса CoS настраиваются заранее (либо вручную на долговременной основе, либо динамически для каждого сеанса связи). При этом формируются таблицы соответствия между классом сервиса CoS и значением кода DSCP (DS code point поля «дифференциальная услуга» DS заголовка IP-пакета [RFC2475]) путем соответствующих настроек параметров в базе данных (БД) граничных маршрутизаторов. На сегодняшний день стандартизовано 14 DSCP. На границе «облака» маршрутизаторов DiffServ потоки агрегируются в группы по значению кода DSCP, который помещается в поле DS заголовка пакета после его классификации. Пакеты, не принадлежащие TCS, обслуживаются в соответствии с механизмом BE.

Основная функциональная нагрузка I-уровня в пакетных ИТС – организация QoS-маршрутизации мультимедийных потоков, что связано с организацией соответствующих дисциплин обслуживания очередей для потоков различных классов в выходных буферах

маршрутизаторов. При этом обслуживание изохронного трафика носит явно выраженный условный характер. Этот факт накладывает определенную специфику на функциональную структуру уровня, связанную с необходимостью «окраски» потоков в сессии. Перенос протокольных блоков (пакетов) многокомпонентных потоков в сессии реализуется в соответствии с таблицами маршрутизации, которые формируются на фазе установления мультимедийного соединения. На I-уровне **функция «совмещения»** реализуется соответствующей «окраской» типа трафика. В **IntServ** на I-уровне «окраска» типа трафика реализуется посредством определенных значений трех старших бит служебного поля «тип сервиса» (Type of Service, TOS) заголовка IP-пакета (подполе приоритета пакета (PRECEDENCE)), а в сети **DiffServ** – значений шести старших битов (код DSCP) поля «дифференциальная услуга» DS заголовка IP-пакета (здесь поле TOS переименовано на DS). Структура и формат заголовка Ipv4, распределение разрядов в поле «тип сервиса» TOS, формат поля «Дифференциальная услуга» и значение кода DSCP приведены в Приложении 2.

Новая версия протокола IPv6 помимо расширения адресного пространства до 128 битовых адресов расширяет и набор опций заголовка, в число которых входят: Hop-by-Hop, Routing (Type 1), Fragment, Destination Option, Authentication, Encapsulation Payload. В IPv6 имеются дополнительные механизмы для «окраски» пакета в сеансе связи. Относительный приоритет доставки пакетов указывается в 8-битовом поле «Класс трафика». Значения приоритетов делятся на два диапазона. Коды от 0 до 7 используются для задания приоритета трафика, для которого отправитель осуществляет контроль перегрузки (например, снижает поток TCP в ответ на сигнал перегрузки). Значения с 8 до 15 используются для определения приоритета трафика, для которого не производится снижения потока в ответ на сигнал перегрузки, например, для пакетов «реального времени».

Практически, уровни приоритета выше или равные 8 зарезервированы для передачи данных в реальном масштабе времени. Для трафика, не критичного к перегрузкам, нижнее значение приоритета (8) должно использоваться для тех пакетов, которые отправитель разрешает выбросить в случае перегрузки (например, видео трафик высокого качества), а высшее значение (15) следует использовать для пакетов, которые отправитель не хотел бы потерять (например, аудио трафик с низкой надежностью). Для «окраски» пакета в сеансе связи может быть использовано и поле заголовка «Метка потока». Поле «Метка потока» позволяет выделять и особым образом обрабатывать отдельные потоки данных без необходимости анализировать содержимое пакетов и может использоваться отправителем для указания на специальную обработку пакета в маршрутизаторе с целью обеспечения требуемого QoS или сервиса реального времени «real-time». «Окраска» потока, пакеты которого требуют соответствующей обработки в маршрутизаторе может быть осуществлена посредством протокола управления или внутри самих пакетов, например, в опции «Hop-by-Hop».

В **физической структуре** сети IntServ функция «совмещения» реализуется в классификаторах пакетов, а в **физической структуре** сети DiffServ – формирователем трафика TC.

В сеансе связи в сети IntServ на каждом транзитном маршрутизаторе, классификатором пакетов организуется контроль предоставления выделенных ресурсов со стороны сети, идентификация и маркировка поступающих пакетов по значениям четырех параметров: IP-адресу и номеру порта источника, IP-адресу и номеру порта приемника (микротоковая идентификация - microflow). Пакеты маркируются в классификаторах посредством определенных значений трех старших бит служебного поля «тип сервиса» TOS заголовка IP-пакета (подполе приоритета пакета (PRECEDENCE)) с целью определения принадлежности определенному потоку и, после прохождения процедуры «политика управления нагрузкой» (policing), направляются в отдельные исходящие очереди для разных потоков на одном и том же порту в соответствии с их приоритетом

или QoS-параметрами. Планировщик пакетов, отвечающий за обработку очередей в соответствии с требуемой дисциплиной обслуживания, осуществляет контроль соответствия параметров входного потока заявленным, распределяет процессорное время и определяет, из какой очереди очередной пакет будет отправлен на обслуживание. Следовательно, им решается задача разделения полосы пропускания исходящего тракта в соответствии с заданными QoS-параметрами для каждого индивидуального потока.

Классификация пакетов в сети DiffServ осуществляется на основании набора параметров «профиль трафика» (traffic profile) из TCS по IP-адресу и номеру выходного порта. При этом определяется соответствие параметров входного потока классу обслуживания CoS, который определяется кодом DSCP в заголовке IP-пакета, т. е. проводится авторизация значения кода в базе данных граничного маршрутизатора. В технологии DiffServ предусмотрено два правила классификации:

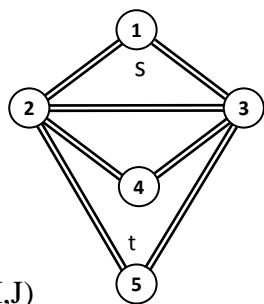
- агрегированное поведение BA (behavior aggregate), когда каждому пакету в результате классификации присваивается агрегированное поведение BA путем присвоения полю DS заголовка пакета соответствующего значения DSCP;
- многопараметрическое MF (multi-field), когда классификация производится на основе информации, содержащейся более чем в одном поле заголовка (например, маркировка пакетов на основе адреса или порта отправителя/получателя, идентификатора протокола, и т. п.).

Агрегированному однотипному потоку в соответствии со значением кодового слова DSCP во всех транзитных маршрутизаторах домена предоставляется одинаковый режим обслуживания РНВ, определяемый классом сервиса CoS. Документ [RFC 2475] определяет режим РНВ как комбинацию функций маршрутизации, классификации, обработки очередей и методов сброса пакетов на каждом шаге передачи пакета от узла к узлу внутри домена DiffServ. Определение необходимого режима обслуживания РНВ в транзитном маршрутизаторе осуществляется формирователем трафика ТС.

5.3. Принцип уровневой иерархии МСС

Пусть задана ЦТМСС, топология которой описывается графом $G = (I, J)$, где I – множество вершин мощности $N = |I|$, $J \subseteq I \times I$ – множество ребер $ij \in J$ рис.5.2. Вершины графа $i \in I$ соответствуют узлам коммутации (УК) сети, в качестве которых могут выступать маршрутизаторы или коммутаторы. Любая пара узлов сети может обмениваться информацией [14]. Не теряя общности, будем полагать, что на вход сети в рамках мультимедийного соединения поступают потоки двух классов: речевой поток класса A или B и интерактивные данные класса C .

Топология сети

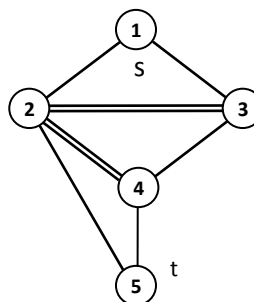


Граф $G(I, J)$

$I = \{1, 2, 3, 4, 5\}$ – вершин
 $J = \{12, 13, 21, 23, 24, 25, 31, 32, 34, 35, 42, 43, 45, 52, 53, 54\}$ – рёбра

$s \in I$ - источник
 $t \in I$ - получатель

Связанный подграф $G_{15} = G(I_{15}, J_{15})$



61 $G_{15} = G(I_{15}, J_{15});$
 $I_{15} = \{1, 2, 3, 4\};$
 $J_{15} = \{12, 13, 23, 24, 25, 32, 34, 42, 45\}$

Рис.5.2 - Построение связного подграфа $G_{15}=G(I_{15},J_{15})$ на графе $G_{15}=G(I_{15},J_{15})$

Таким образом, в рамках единой сессии перенос мультимедийных объектов связан с функционированием двух приложений, которые порождают в совокупности многокомпонентный информационный поток. При этом будем полагать, что каждая компонента этого потока функционирует в сессии независимо друг от друга, но ограничена ее длительностью. В графе $G=(I,J)$ выделяются две вершины $s \in I$ – источник и $t \in I$ – получатель и на нем строится связный подграф $G_{st}=(I_{st},J_{st})$. Объемные величины входных мультимедийных потоков заданы в виде матриц $\widehat{Y}^{multy} = \|\widehat{a}_{st}^{multy}\|$ (Эрл). Здесь $\widehat{Y}^{multy} = Y^{multy}(1-b^{multy})$ – пропущенная мультимедийная нагрузка, b^{multy} – заданная вероятность потерь мультимедийного вызова, $\widehat{a}_{st}^{multy} = a_{st}^{multy}(1-b_{st}^{multy})$ – мультимедийная нагрузка, пропущенная в тракт $st \in S$. Все пары st , для которых $a_{st}^k \neq 0$, образуют множество корреспондирующих пар S^k мощности $|\Omega| = |S^k|$. Каждое ребро графа $ij \in J$ моделирует межузловой линейно-цифровой тракт (ЛЦТ) связи и ему поставлена в соответствие величина V_{ij} – скорость передачи (бит/с) и P_{ij} – вероятность ошибки в тракте.

Считаем, что топология сети не изменяется и входные потоки стационарны, для каждой пары $st \in S^k$ определено несколько статических путей в подграфе $G_{st}=(I_{st},J_{st})$, необязательно совпадающих для различных потоков мультимедийного соединения. Мультимедийный трафик между трактами $st \in S^k$ в рамках единой транспортной услуги может быть распределен в сессии по нескольким путям в фиксированных во времени определенных пропорциях.

Введем следующие обозначения: $\widehat{l}_{st,m}^k = \{s, i_1, i_2, \dots, i_{p-1}, t\}_{st,m}^k$ – путь m -го выбора ($m=1, M_{st}^k$) длины p для пары $st \in S^k$ с упорядоченными ребрами относительно источника $s \in I_{st}$. Корневое дерево $R_{st}^k = \{\widehat{l}_{st,m}^k, m=1, M_{st}^k\}$, которое в общем случае может быть поддеревом дерева всех путей из s в t . Множество ребер $l_{st,m}^k = \{ij \in J : ij \in \widehat{l}_{st,m}^k\}$ моделирует путь m -го выбора $\widehat{l}_{st,m}^k$ мощности $r_{st,m}^k = |M_{st,m}^k|$. В общем случае $\bigcap_{m=1}^{M_{st}^k} l_{st,m}^k \neq \emptyset$.

Будем считать, что маршруты передачи изохронного трафика независимы.

Корневое дерево путей R_{st}^k лежит в основе построения последовательно-параллельной схемы путей L_{st}^k , моделирующих маршруты передачи мультимедийных потоков в сессии. Построение схемы мультимедийных соединений осуществляется с помощью «расклеивания» общих транзитных узлов дерева R_{st}^k по разным путям L_{st}^k и «склеивания» висячих вершин $t \in I_{st}$ в один узел. Каждый пакет при его вводе в сеть в сессии с глобальной вероятностью $p_{st,m}^k$ направляется в m -й путь, выбранный на фазе установления соединения из множества L_{st}^k . $\sum_{m=1}^{M_{st}^k} p_{st,m}^k = 1$ для $\forall st \in S^k$ (рис. 5.3).

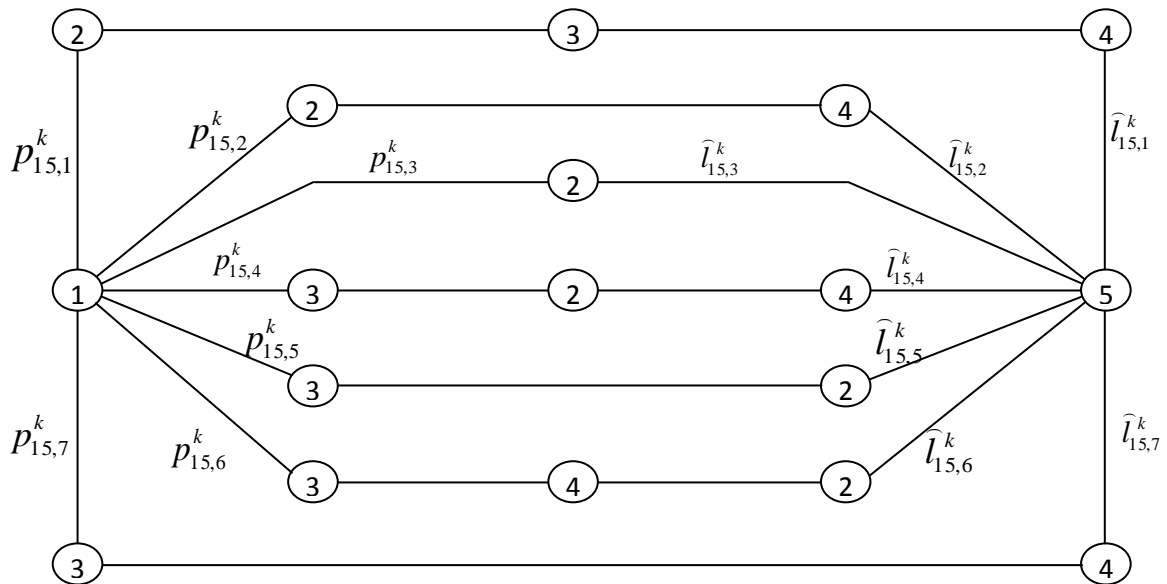


Рис. 5.3 - Пример организации параллельно-последовательных схем L_{15}^k на связном подграфе $G_{15} = G(I_{15}, J_{15})$

Такая процедура маршрутизации является частным случаем общей процедуры, когда в каждом УК для пакета между парой $st \in S^k$ выбирается исходящий тракт ЛЦТ $ij \in J$ с локальной вероятностью p_{ij}^k . При этом $p_{st,m}^k = \prod_{ij \in I_{st,m}^k} p_{ij}^k$ для $\forall st \in S^k$.

Функционирование ИТС в виду их сложности можно описать комплексом моделей, отражающих работу иерархически разделенных сетевых функций, сгруппированных в логически завершённые функциональные группы задач (логические уровни или логические подсистемы) ее архитектуры. При построения моделей функционирования ИТС для учета вертикальной взаимообусловленности и соподчиненности уровневой организации их логических структур предлагается использовать принцип иерархии, который *определяет* тип структурных отношений в сложных многоуровневых системах, характеризуемых упорядоченностью, организованностью взаимодействий между отдельными уровнями по вертикали, позволяющий разложить систему по парам «отправитель-получатель» и учесть в рамках единой модели всю протокольную вертикаль логических соединений, поддерживающих мультимедийную сессию.

В рамках этого подхода эффективность функционирования ИТС в режиме установленного соединения предлагается оценивать с помощью набора уровневых функционалов $K_{h,ij}^{k,q}$ оценки эффективности использования пропускной способности системы трафиком k -го класса q -го типа, которые определяют требуемую долю пропускной способности V_{ij} линейно-цифрового тракта (ЛЦТ) для организации мультимедийного соединения (здесь индексами обозначены: h - соответствующий уровень архитектуры ИТС, т. е. на транспортном уровне сеть выступает как набор трактов переноса многокомпонентных потоков в рамках единой мультимедийной сессии. При этом указанные функционалы (в отличие от [9.37]) зависят не только от необходимой для их работы служебной информации соответствующих объемов, но и от протоколов функционирования отдельных уровней архитектуры ИТС, поддерживающих соответствующие службы. Ясно, что при их построении необходимо учитывать также, что использование ЛЦТ пакетами данных зависит от их загрузки речевыми пакетами, обслуживаемых в системе с абсолютным приоритетом. Поэтому критерий эффективности

K_{ij}^C использования ЛЦТ пакетами данных будет зависеть от параметров критерия эффективности K_{ij}^B использования ЛЦТ речевыми пакетами, т. е. носит выраженный условный характер. С учетом указанных предположений оптимизацию предложенных функционалов можно проводить поэтапно, оптимизируя сначала использование ЛЦТ речевым трафиком, а затем вычисляя максимум функционала K_{ij}^C использования ЛЦТ трафиком данных при условии, что параметры K_{ij}^B оптимальны и фиксированы. Согласно логической структуре ИТС комплексные функционалы K^k должны «расслоиться» на отдельные уровневые функционалы. Аналогичный подход был предложен для построения комплексного критерия эффективности ТС цифровых сетей с интеграцией служб [9.38] и апробирован также при построении комплексных критериев эффективности ИТС на технологиях IP-QoS и ATM [9.11, 9.19, 9.38–9.42].

В силу относительной независимости уровней архитектуры и вложенности протокольных блоков (уровневых примитивов) верхних уровней архитектуры ИТС в информационное поле протокольных блоков нижележащих уровней функционалы $K_{h,ij}^k$ обладают мультипликативным свойством. В этой связи критерий эффективности использования пропускной способности ЛЦТ трафиком k -го класса q -го типа

$$K_{ij}^k = \prod_h K_{h,ij}^k, \quad (5.1)$$

На примере эталонной модели ВОС заметим, что в $h=1$ соответствует физическому Ph -уровню, $h=2$ – каналному L -уровню, а $h=3$ – сетевому N -уровню. В (5.1) учитываются только уровни, расположенные ниже сетевого, так как протоколы T -уровня функционируют для транспортных соединений определенного класса трафика, включающих совокупность трактов маршрута.

Обозначим через $K_{st,m}^k$ – критерий эффективности использования m -го виртуального пути из множества $L_{st}^{k_p}$ трафиком k -го класса. Этот функционал можно трактовать как коэффициент передачи системы, составленной из цепочки каналов $ij \in l_{st,m}^k$ заданной пропускной способности. С учетом чистой ретрансляции пакетов через указанную систему (на выходе системы ослабления входного сигнала не происходит) $K_{st,m}^k$ можно представить в виде среднегеометрического составляющих его критериев эффективности ЛЦТ K_{ij}^k [9.43]:

$$K_{st,m}^k = K_T^k r_{st,m} \sqrt{\prod_{ij \in l_{st,m}^k} K_{ij}^k}, \quad (5.2)$$

где K_T^k – критерий, учитывающий эффективность использования пропускной способности ЛЦТ трафиком k -го класса на транспортном уровне.

В силу того, что транспортное виртуальное соединение может быть организовано между парой $st \in S^k$ по нескольким виртуальным путям, выражение для общего критерия эффективности использования всех транспортных соединений трафиком k -го класса имеет следующий вид:

$$K_{st}^k = \sum_{m=1}^{M_{st}^k} p_{st,m}^k K_T^k r_{st,m} \sqrt{\prod_{ij \in l_{st,m}^k} K_{ij}^k}, \quad (5.3)$$

С учетом того, что весовой коэффициент соединений на транспортном уровне между парой $st \in S^k$ определяется долей входящего потока k -го типа \widehat{a}_{st}^k в общем входящем сетевом потоке в рамках всей поступающей мультимедийной сетевой нагрузки $Y^{multy} = \sum_{st \in S^k} \widehat{a}_{st}^k$, сетевой критерий эффективности использования пропускной способности системы трафиком k -го класса имеет вид

$$K^k = r^k \sqrt{\sum_{st \in S^k} \frac{\widehat{a}_{st}^k}{Y^{multy}} (K_{st}^k)^2}, \quad r^k = |S^k|, \quad (5.4)$$

Функционал подобного типа обладает свойством $\lim_{q \rightarrow \infty} K^k = \max_{st \in S^k} K_{st}^k$, что не дает явного преимущества транспортным соединениям с большими весовыми коэффициентами [9.43].

Построенный критерий эффективности ИТС отображает основное назначение системы исходя из цели проектирования и позволяет произвести его оценку в количественной и численной форме в режиме установленного соединения. При этом он имеет четкий физический смысл, чувствителен к основным варьируемым параметрам ИТС, что позволяет определить наиболее рациональные их характеристики, а также обеспечивает сравнение различных вариантов построения для выбора наиболее рационального при заданных условиях проектирования. Выражение (5.5) в общем виде представляет собой системную модель инфотелекоммуникационной системы.

Модели логических мультимедийных соединений, используя принцип иерархии, могут быть представлены в виде функциональных зависимостей требуемой пропускной способности системы для переноса мультимедийного трафика от соответствующих критериев эффективности ИТС [9.11, 9.15, 9.19]. Так, например, модели уровневых логических соединений k -го класса даются выражением

$$V_{h,ij}^{\min k} = V_{ij} K_{h,ij}^k, \quad (5.5)$$

5.4. Формализация критериев эффективности ИТС-IP-QoS

5.4.1. Критерий эффективности транспортного уровня

Согласно модели архитектуры ИТС на технологии IP-QoS комплексный критерий эффективности ИТС K_{ij}^k должен «расслоиться» на уровневые критерии $K_{h,ij}^k$ h -х уровней ее логической структуры: транспортного, межсетевого и сетевых интерфейсов. Здесь $h=1$ соответствует физическому уровню, $h=2$ – уровню сетевого интерфейса, $h=3$ – уровню межсетевого взаимодействия IP, а $h=4$ – транспортному уровню уточненной логической структуры архитектуры DARPA-QoS.

В сессии фрагменты активной речи, видео- и/или сообщения данных поступают на транспортный уровень сетевых драйверов мультимедийных терминалов и/или серверов инфоуслуг сети. При поступлении информационных массивов в порт маршрутизатора и/или оконечной системы программы управления передачей транспортного уровня формирует соответствующие протокольные блоки (сегменты) из неструктурированного потока трафика класса B и данных класса C . При установлении транспортного канала пользователь обслуживается через конкретный логический порт. Введем следующие дополнительные обозначения: β^C – коэффициент, учитывающий механизм организации обратной связи на транспортном уровне (для протокола TCP) с целью защиты от ошибок блоков данных класса C ; H_{TCP} – длина заголовка TCP-блока данных, бит; H_{UDP} –

длина заголовка *UDP*-блока данных, бит; H_{IP} – длина IP-заголовка, бит; H_{NA} – длина заголовка протокольного блока сетевого интерфейса бит; τ^B – средняя длительность активного фрагмента речевого трафика *B*-класса, с; s^C – средняя длина сообщения данных на транспортном уровне, включая связной заголовок транспортного уровня, бит; N^B – среднее число информационных частей речевого пакета в активном речевом фрагменте на транспортном уровне; N^C – среднее число информационных частей пакета данных в сегменте данных на транспортном уровне; L^B – длина пакета речевого трафика *B*-класса, бит; L^C – длина пакета данных, бит.

Плотность «набивки» полезной нагрузкой информационных полей протокольных блоков транспортного уровня учитывают соответствующие коэффициенты использования пропускной способности ЛЦТ на этих подуровнях:

$$\frac{s^B}{N^B(L^B - H_{IP})}; \frac{s^C}{N^C(L^C - H_{IP})}.$$

Критерий эффективности транспортного уровня K_{TCP}^C учитывает также особенность работы протокола *TCP* при организации защиты от ошибок трафика класса *C*, на которую также затрачивается часть пропускной способности ЛЦТ. Эти потери оцениваются множителем β^C . Отметим, что чем длиннее нарезаемая порция информации, тем больше в среднем передается отрывков пауз между речевыми фрагментами или сообщениями данных. При уменьшении этой порции эффективность использования ЛЦТ передачи снижается за счет увеличения служебной избыточности, переносимой по каналам связи в заголовках протокольных блоков уровня.

Таким образом, с учетом введенных предположений выражения для уровневых функционалов K_{TCP}^C и K_{UDP}^B транспортных логических соединений имеют следующий вид (протокольной избыточностью, вносимую заголовками уровневых примитивов в *C*-сообщении и *B*-речевом фрагменте, мы пренебрегаем):

$$K_{UDP}^B = \frac{s^B}{N^B(L^B - H_{IP})}, \quad .. (5.6)$$

$$K_{TCP}^C = \frac{s^C}{N^C(L^C - H_{IP})} \beta^C. \quad (5.7)$$

5.4.2. Критерий эффективности межсетевого уровня

Поддержка мультимедийного логического транспортного соединения может осуществляться либо несколькими логическими соединениями межсетевого уровня для каждого класса трафика, либо одним мультиплексным логическим соединением в определенном классе, поддерживающим несколько типов соединений.

Для формализации логических соединений межсетевого уровня введем следующие обозначения: ρ_{ij}^B и ρ_{ij}^C - коэффициенты загрузки межсетевого уровня соответственно речевыми *B*-пакетами и пакетами данных класса *C*. Указанные коэффициенты учитывают динамику очередей разнородных пакетов на межсетевом уровне, так как даже наличие абсолютного приоритета для речевых пакетов не исключает очередей пакетов указанного класса трафика на маршрутизаторе, принадлежащих различным речевым соединениям. Часть пропускной способности ЛЦТ, которая тратится на служебную информацию, вводимую межсетевым уровнем для всех классов трафика, можно оценить

так: $\frac{L^k - H_{IP}}{L^k}$. Таким образом, критерии эффективности использования пропускной способности ЛЦТ разнородным трафиком на межсетевом уровне с учетом коэффициентов загрузки речевыми пакетами ρ_{ij}^B и пакетами данных ρ_{ij}^C , которые учитывают динамику очередей на этом уровне (через указанные коэффициенты в задачах анализа сети учитывается интенсивность потока поступления пакетов в транспортный канал и интенсивность их обслуживания) соответственно

$$K_{ij,IP}^B = \frac{L^B - H_{IP}}{L^B} \rho_{ij}^B, \quad (5.8)$$

$$K_{ij,IP}^C = \frac{L^C - H_{IP}}{L^C} \rho_{ij}^C \quad (5.9)$$

5.4.3. Критерий эффективности уровня сетевого доступа

Критерии эффективности использования пропускной способности ЛЦТ разнородным трафиком на уровне сетевых интерфейсов имеют вид

$$K_{ij,NA}^B = \frac{L^B}{L^B + H_{NA}}, \quad (5.10)$$

$$K_{ij,NA}^C = \frac{L^C}{L^C + H_{NA}}. \quad (5.11)$$

Не нарушая общности рассуждений в дальнейшем будем полагать, что для Ph-уровня $K_{ij,1}^{B,q} = K_{ij,1}^C \cong 1$. Формирование уровневых функционалов использования пропускной способности ИТС в архитектуре *DARPA* приведен в таблице 5.1.

Таблица 5.1 - Формирование уровневых функционалов использования пропускной способности ИТС для незащищенных логических соединений в уточненной архитектуре *DARPA*

Уровень архитектуры DRPA	Структура протокольного блока	Уровневый функционал	
		Обозначение	Структура
Верхний	s^B, s^C сообщение, речевой фрагмент		
Транспортный (TCP)	$L^k - H_{IP}$	K_{UDP}^B, K_{TCP}^C	$\frac{s^B}{N^B(L^B - H_{IP})}, \frac{s^C}{N^C(L - H_{IP})} \beta^C$
Межсетевого взаимодействия (IP)	L^k -----,---- $L^k - H_{IP}, H_{IP}$	$K_{ij,IP}^B, K_{ij,IP}^C$	$\frac{L^B - H_{IP}}{L^B} \rho^B; \frac{L^C - H_{IP}}{L^C} \rho^C$
Сетевого интерфейса (NA)	$L^k + H_{NA}$ -----,-- L^k, H_{NA}	$K_{ij,NA}^B, K_{ij,NA}^C$	$\frac{L^B}{L^B + H_{NA}}; \frac{L^C}{L^C + H_{NA}}$

Примечание. В таблице 5.1 -----,---- (обозначение протокольного блока уровня, включающий сервисный блок данных вышележащего уровня и управляющую служебную информацию уровня –заголовок).

5.4.4. Комплексный критерий эффективности инфотелекоммуникационных транспортных систем на технологии IP-QoS

С учетом выше приведенных формул выражения для комплексных критериев эффективности использования ЛЦТ соответственно трафиками классов B и C в режиме установленного соединения в ИТС-П на технологии IP-QoS:

$$K_{ij}^B = \frac{s^B \rho_{ij}^B(L^B)}{N^B(L^B + H_{NA})}, \quad (9.12)$$

$$K_{ij}^C = \frac{s^C \rho_{ij}^C(L^B, L^C, \rho^B)}{N^C(L^C + H_{NA})}, \quad (9.13)$$

Видно, что критерии эффективности использования ЛЦТ для данных уже на двух нижних уровнях архитектуры *DARPA* носит условный характер, так как величины ρ_{ij}^C зависят от длин речевых пакетов и удельной загрузки ЛЦТ речевыми B -пакетами; T_{ij}^C – заданное среднее время пребывания C -пакета в ЛЦТ.

Выражение для общих критериев эффективности K_{st}^B и K_{st}^C использования транспортных соединений трафиками классов B и C имеют следующий вид:

$$K_{st}^B = \sum_{m=1}^{M_{st}^B} p_{st,m}^B \frac{s^B}{N^B(L^B + H_{NI})} {}_{r_{st,m}^B} \sqrt{\prod_{ij \in I_{st,m}^B} \rho_{ij}^B(L^B)},$$

$$K_{st}^C = \sum_{m=1}^{M_{st}^C} p_{st,m}^C \frac{s^C \beta_{ij}^C(L^C, p_{ou})}{N^C(L^C + H_{NI})} {}_{r_{st,m}^C} \sqrt{\prod_{ij \in I_{st,m}^C} \rho_{ij}^C(L^B, \rho_{ij}^B, L^C)}. \quad (5.14)$$

Общесетевые комплексные критерии эффективности использования пропускной способности пакетной ИТС произвольной топологии трафиками классов B и C даются соответственно выражениями

$$K^B = {}_{r^{Ba}} \sqrt{\sum_{st \in S^{Ba}} \frac{\hat{a}_{st}^B}{Y^{multy}} (K_{st}^B) r^B}, \quad K^C = {}_{r^{Bq}} \sqrt{\sum_{st \in S^C} \frac{\hat{a}_{st}^C}{Y^{multy}} (K_{st}^C) r^C}. \quad (5.15)$$

Функциональные зависимости (5.15) представляют собой системную формализацию процессов переноса мультимедийного трафика в ИТС IP-QoS в рамках предоставления инфокоммуникационной услуги связи с заданными QoS-нормами в режиме установленного соединения. В (5.15) \hat{a}_{st}^k – пропущенная нагрузка k -го класса в направлении $st \in S^k$; $Y^{multy} = \sum_k Y^k$; $Y^k = \sum_{S^k} \hat{a}_{st}^k$; $\hat{a}_{st}^k = \frac{Mark^k}{3600c} y^{multy} \cdot t^{multy} b_{st}$ (Эрл), y^{multy} -

интенсивность мультимедийных вызовов (вызов/час); t^{multy} – средняя длительность мультимедийного соединения, с; b_{st} - потери мультимедийного вызова на направлении $st \in S^k$; Θ^{\min} - базовая минимальная ширина полосы пропускания (бит/с);

$$Mark^k = \sum_{q=1}^Q \left[\frac{v_q^k}{\Theta_{\min}} \right]$$
 - суммарная марка трафика, требуемых для обслуживания всех пользователей k -го класса в тракте $ij \in J$; v^k - скорость работы окончного устройства (бит/с); Q - общее количество мультимедийных окончных устройств.

5.5. Аналитические модели процессов функционирования однородной инфотелекоммуникационной транспортной IP-QoS-системы

Аналитические модели процессов функционирования ИТС строятся в виде набора моделей уровневых логических соединений для переноса мультимедийного трафика, отличающиеся учетом сложной многокомпонентной информационной потоковой структуры мультимедийных соединений и специфики их обслуживания.

5.5.1. Аналитическое описание транспортных логических соединений IP-QoS-ИТС

В режиме установленного виртуального мультимедийного соединения фрагменты речевого трафика и сообщения данных поступают на транспортный уровень от окончных мультимедийных терминалов. Речевой блок (РБ) – элементарная порция речевой дискретной информации, вырабатываемая вокодером или липледером в промежутке между двумя соседними синхроимпульсами. Величина РБ зависит от скорости передачи и алгоритма преобразования речевого сигнала в дискретную форму. Например, для вокодеров, работающих по принципу измерения мгновенного спектра речевого сигнала (полосовые вокодеры), при скорости 2400 бит/с – порядка 50 бит; для липледеров на 9,6 кбит/с величина РБ равна 260 бит. Последовательность РБ совместно с сигналом, формируемым детектором пауз, образует речевой сегмент (РС) – своеобразное «сообщение» длиной s^B в диалоговом речевом соединении, т. е. под речевым сегментом понимается отрезок речевого сигнала, состоящего из системы активных звуковых сегментов, чередующихся с паузами, длина которых меньше или равна наперед заданной величины. Эту величину назовем формирователем РС (ФРС). Известно [18], что основное влияние на качество субъективного восприятия речевых сигналов, прошедших через цифровую систему связи, оказывают следующие два фактора: длина речевого сегмента (РС) и сохранение непрерывности РС. Речевой сегмент преобразуется и кодируется в цифровой системе как единое целое, образуя своеобразное «сообщение». Длина РС не настолько коротка, чтобы содержать только одно слово или слог, но и не настолько длинна, чтобы содержать одну фразу или изречение. Обычно длина РС выбирается таким образом, чтобы при его обработке и транспортировке сохранялась непрерывность звучания звуковых сегментов на приемном конце ИТС. Как показывают эксперименты, влияние переменной составляющей сетевой задержки наиболее ощутимо при наличии коротких пауз между словами по сравнению с более длинными паузами между фразами и изречениями. Отсюда следует, что набор длины РС необходимо производить с учетом статистики речевых сигналов [18].

Пакеты формируются из соответствующих фрагментов и сообщений s^{kp} . Можно показать [14], что для любых $F^{Bp}(t)$ и $F^{Cp}(l)$:

$$\frac{s^{Bp}}{N^{Bp}} \rightarrow L^{Bp} - H_{IP} \quad \text{и} \quad \frac{s^{Cp}}{N^{Cp}} \rightarrow L^{Cp} - H_{IP} \quad \dots\dots\dots (5.16)$$

Этот факт существенно упрощает процедуры вычислений функционалов и может применяться на ранних этапах проектирования сети для оценочных расчетов в условиях недостаточности данных.

В предположении, что длительности речевых фрагментов и длины сообщений распределены экспоненциально с параметрами $1/\tau^B$ и $1/s^k$, то можно показать, что

$$\frac{s^{Bp}}{N^{Bp}} = v^B \tau^B \left(1 - e^{-\frac{L^{Bp} - H_{IP}}{v^B \tau^B}}\right) \text{ и } \frac{s^{Cp}}{N^{Cp}} = s^{Cp} \left(1 - e^{-\frac{L^{Cp} - H_{IP}}{s^{Cp}}}\right),$$

где $\tau^B = \int_0^{\infty} t dF^{Bp}(t)$ и $s^{Cp} = \int_0^{\infty} l dF^{Cp}(l)$. (5.17)

Примером, когда условие $s^B \rightarrow \infty$ можно считать выполненным, является пакетизатор с кодированием пауз. В этом случае поток речевых пакетов в виртуальном соединении образует единое сообщение, длина которого равна длительности сеанса, умноженной на скорость работы речепреобразующего устройства.

Влияние процедур организации обратной связи для повторной передачи поврежденного кадра, обнаруженного на приемном конце составного тракта передачи, оценивается на транспортном уровне коэффициентом β_{st}^k [14]. Выше отмечалось, что для речевых пакетов недопустимы переспросы, но могут допускаться их определенные потери. Поэтому в дальнейшем будем считать $\beta_{st}^B = 1$. Величина β_{st}^{Cp} является функцией длины кадра $L^{Cp} + H_{NA}$ и вероятности ошибки в составном тракте p_{ou} . Для упрощения вычислений примем, что уровень ошибок во всех каналах сети одинаков, т. е. $\beta_{st}^{Cp} = \beta^C(L^{Cp}, p_{ou})$. Если обозначить p_0 вероятность отсутствия ошибок в кадре длины $L^{Cp} + H_{NA}$ и предположить, что число переспрашиваемых пакетов подчинено геометрическому распределению, то для РОС [14]

$$\beta^{Cp} = \sum_{k=1}^{\infty} \frac{1}{k} p_0 (1 - p_0)^{k-1} = -\frac{p_0}{1 - p_0} \ln p_0. \quad \dots\dots\dots (5.18)$$

В частности, для биномиального канала с вероятностью ошибки, равной p_{ou} , $p_0 = (1 - p_{ou})^{L^{Cp} + H_{NA}}$. Для каналов с группирующимися ошибками выражение для p_0 может быть получено, например, из модели [26]. Вывод формулы (5.18) приведен в [14]. Указанный коэффициент должен быть учтен в общей модели виртуального соединения для переноса трафика класса C в составном тракте $st \in S^{Cp}$.

В результате, критерии эффективности межузловых логического соединений на транспортном уровне в целях достижения минимальной пропускной способности для переноса изохронного трафика класса B и трафика данных класса C имеют вид

$$K_{ij, TCP}^{Bp} = \frac{v^B \int_0^{\infty} t dF^{Bp}(t)}{\sum_{k=1}^{\infty} k \left[F^{Bp}\left(\frac{L^{Bp} - H_{IP}}{v^{B,q}} k\right) - F^{Bp}\left(\frac{L^{Bp} - H_{IP}}{v^{Bp}} (k-1)\right) \right] (L^{Bp} - H_{IP})}, \quad (5.19)$$

$$K_{ij, TCP}^{Cp} = \frac{\int_0^{\infty} l dF^{Cp}(l)}{\sum_{k=1}^{\infty} k \left[F^{Cp}((L^{Cp} - H_{IP})k) - F^{Cp}((L^{Cp} - H_{IP})(k-1)) \right] (L^{Cp} - H_{IP})} \left(-\frac{p_0}{1 - p_0} \ln p_0 \right) \quad (5.20)$$

5.5.2. Аналитическое описание межсетевых логических соединений

Важнейшей частью процесса обработки пакетов является алгоритм управления очередью в буферах сетевого оборудования. Базовой идеей принципа «управления буферным пространством» является классификация пакетов и соответствующий ей сброс пакета или помещение его в буфер. В свою очередь, введение дополнительных буферов в структуру маршрутизатора расширило возможности реализации принципа «справедливого распределения ресурсов»: стало возможным управлять доступом пакетов определенных потоков к процессорному времени маршрутизатора. На основе этих двух базовых идей были разработаны методы, названные соответственно Queue Management (управление очередью) и Scheduling (планирование обслуживания). Под алгоритмом управления очередью понимается набор методов, управляющих поступлением, хранением и передачей на обслуживание поступающих в систему пакетов. Основные алгоритмы организации очередей – это алгоритм отбрасывания «хвоста» при заполнении максимального размера очереди (Tail Drop) и алгоритм раннего обнаружения (Random Early Detection, RED), когда прибывший пакет отбрасывается с вероятностью, зависящей от того, насколько превышен установленный порог или его модификация – алгоритм взвешенного раннего обнаружения переполнений (WRED), который используют приоритеты при выработке политики предотвращения переполнения выделенного буферного пространства [27].

Планирование обслуживания в ИТС базируется на введении приоритетного выбора очередей. Аналитические модели логических соединений на межсетевом уровне с учетом предположений и ограничений раздела 2, в общем случае описывающие процесс передачи пакетов класса B и пакетов данных в системе в рамках предоставления инфокоммуникационной услуги, может быть представлена однолинейной многофазной системой массового обслуживания (СМО) с ожиданием и абсолютным приоритетным обслуживанием (с дообслуживанием) B -пакетов по отношению к пакетам данных на каждой фазе. Внутри класса B могут быть организовано либо раздельное обслуживание различных типов трафика с относительным приоритетом обслуживания индивидуальных очередей, либо обслуживание всех типов трафика в режиме FIFO в одной очереди с равным приоритетом.

Для однородных пакетных ИТС расчет сетевых характеристик логического соединения уровня межсетевого взаимодействия сводится к расчету параметров «типичного» для сети n -звенного составного канала. Математическая модель, описывающая процесс передачи трафика классов B и C по однородному транспортному каналу, может быть представлена однолинейной многофазной СМО с ожиданием и абсолютным приоритетным обслуживанием (с дообслуживанием) речевых пакетов по отношению к пакетам данных на каждой фазе). Каждый канал $ij \in J$ ЛЦТ совместно с соответствующей ему частью памяти маршрутизаторов моделируется СМО $M/M/1/\infty$. Внутри класса B могут быть организовано либо раздельное обслуживание различных типов трафика с относительным приоритетом обслуживания индивидуальных очередей, либо обслуживание всех типов трафика в режиме FIFO в одной очереди с равным приоритетом.

Для вывода выражений коэффициентов загрузки ρ_{ij}^{Bp} и ρ_{ij}^{Cp} межсетевого уровня соответственно речевыми B -ячейками и ячейками данных класса C , которые учитывают динамику очередей разнородных пакетов на межсетевом уровне, сделаем следующие допущения и предположения:

- суммарные потоки на входе каждого канала тракта независимы друг от друга и являются простейшими;

- допустимое время сквозной задержки речевых ячеек в транспортном канале θ_{st}^{Bp} включает в себя только две основные компоненты: время накопления информационной

части в оконечной системе источника, связанное с накоплением информационного поля ячейки в речепреобразующем устройстве (на передаче $T_{pak}^B = (L^{Bp} - H_{IP})/v^B$ и время задержки в тракте передачи T_{st}^{Bp} . Последнее включает случайную задержку пакетов T_i^{Bp} в маршрутизаторах, обусловленную а) обслуживанием пакетов в поле коммутации, связанным с анализом таблиц маршрутизации и обработкой ячеек в коммутационных элементах (зависит от величины суммарного входящего потока и производительности связанных процессоров); б) обслуживанием планировщиком пакетов T_{ij}^{Bp} (определяется объемом выходного буфера, дисциплиной обслуживания исходящей очереди и скоростью V_{ij} передачи в ЛЦТ). Заданное среднее время пребывания в тракте передачи ячеек класса C T^{Cp} также содержит время накопления информационной части у абонента, равное $T_{pak}^{Cp} = (L^{Cp} - H_{IP})/\omega^C$ и T_{st}^{Cp} . Отсюда возникает ограничение на величины $L^{Bp} \triangleleft \theta_{st}^{Bp} v^B - H_{IP}$ и $L^{Cp} \triangleleft T_{st}^{Cp} \omega^C - H_{IP}$, которое необходимо учитывать при решении задачи параметрического анализа сети;

- в общем балансе времени θ_{st}^{Bp} и T^C мы не будем учитывать время, затраченное на «последней миле» или в сети абонентского доступа T_{lm} , задержку распространения сигнала T_{pf} в физической среде и задержку T_{jitter}^B на стороне получателя в депакетизаторе, связанной с компенсацией величины флуктуации сетевой задержки для обеспечения непрерывного воспроизведения речевых фрагментов на приеме.

На рис.5.4 представлена модель однородного логического соединения ИТС-IP-QoS на уровне межсетевое соединения в режиме установленного соединения.

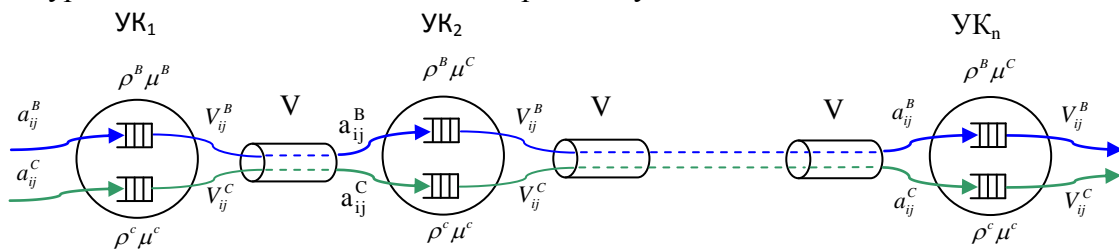


Рисунок 5.4. Модель однородного логического соединения уровня межсетевое соединения в ИТС-IP-QoS

Транспортный канал представляет собой n последовательно включенных межузловых ЛЦТ от источника до получателя (этот параметр учитывает качество маршрутизации информации в сети и ее топологию). На физическом уровне ЛЦТ образуют составной тракт передачи. При сделанных допущениях определим выражения для коэффициентов загрузки ρ_{ij}^{Bp} , ρ_{ij}^{Cp} . Максимальный коэффициент загрузки тракта ρ_{ij}^{Bp} пакетами трафика класса B можно определить из следующих соображений. Пусть t_i - время пребывания пакета в одном ЛЦТ составного VBRrt-канала передачи, тогда общее время пребывания B -пакета в n -звеном канале передачи есть $T_{st}^{Bp} = \sum_{i=1}^n t_i$.

Известно, что преобразование Лапласа плотности распределения времени пребывания B -пакета в n -звеном составном однородном канале, каждый ЛЦТ которого моделируется СМО типа $M/M/1$ для любой пары $st \in S^{Bp}$ имеет вид

$f_{st}^{Bp}(s) = \prod_{i=1}^n f_i^{Bp}(s) = \prod_{i=1}^n \frac{\mu_i^{Bp}(1-\rho_i^{Bp})}{s + \mu_i^{Bp}(1-\rho_i^{Bp})}$. В частности, если тракт передачи однороден, то $f_i^{Bp}(s) = f^{Bp}(s)$ для $\forall i$ и

$$f_{st}^{Bp}(s) = \left[\frac{\mu^{Bp}(1-\rho^{Bp})}{s + \mu^{Bp}(1-\rho^{Bp})} \right]^n, \quad (5.21)$$

где $\mu^{Bp} = \frac{V}{L^{Bp} + H_{IP}}$ - величина, обратная средней длительности обслуживания речевой ячейки в i -ой СМО типа $M/M/1$ ($i = \overline{1, n}$); $\rho_{ij}^{Bp} = \rho^{Bp}$ - максимальная загрузка этой системы пропущенным речевым трафиком. Обращение преобразования (5.21) дает плотность распределения вероятности $f_{st}^{Bp}(t)$ времени пребывания B -пакета в сквозном транспортном канале: $f_{st}^{Bp}(t) = \frac{[\mu^{Bp}(1-\rho^{Bp})]^n t^{n-1}}{(n-1)!} e^{-\mu^{Bp}(1-\rho^{Bp})t}$. Отсюда получаем уравнение для нахождения максимально допустимой величины ρ^{Bp} :

$$\frac{[\mu^{Bp}(1-\rho^{Bp})]^n}{(n-1)!} \int_{\theta_{st}^{Bp}}^{\infty} t^{n-1} e^{-\mu^{Bp}(1-\rho^{Bp})t} dt = d^B, \quad (5.22)$$

Обозначив $x = \mu^{Bp}(1-\rho^{Bp})t$ и произведя замену переменных в уравнении (5.22), получим

$$\int_z^{\infty} x^{n-1} e^{-x} dx = (n-1)! d^B, \quad (5.33)$$

Здесь

$$z = \zeta(1-\rho^{Bp}), \quad (5.34)$$

где $\zeta = \left[\frac{\theta_{st}^B V}{L^{Bp} + H_{NA}} - \frac{L^{Bp} - H_{IP}}{L^{Bp} + H_{NA}} \frac{V}{v^B} \right]$. По определению левая часть уравнения (5.33) есть

неполная гамма-функция $\Gamma(n, z) \equiv \int_z^{\infty} x^{n-1} e^{-x} dx$, следовательно

$$\Gamma(n, z) = (n-1)! d^B. \quad (5.35)$$

По аргументу n справедливо следующее функциональное уравнение

$$\Gamma(n+1, z) = n\Gamma(n, z) + z^n e^{-z}. \quad (5.36)$$

Для любого n уравнение (5.35) с учетом рекуррентного соотношения (5.36) имеет вид:

$$\frac{1}{(n-1)!} z^{n-1} + \frac{1}{(n-2)!} z^{n-2} + \dots + \frac{1}{1!} z + 1 = d^B e^z, \quad (5.37)$$

а условие существования его решения при условии $0 \leq \rho^B \leq 1$ есть [11]

$$-\ln d^B \leq \zeta - \ln \sum_{k=0}^{n-1} \frac{\zeta^k}{k!} \quad (5.38)$$

Если это условие не выполняется, то при заданных d^B, n и ζ трафик класса B принципиально не может быть обслужен.

Левая часть уравнения (5.37) есть полином степени $n-1$ относительно переменной z . Для решения приведенного выше трансцендентного уравнения (5.37) может быть построена итерационная процедура, для чего прологарифмировав это уравнение и разрешив его относительно z , получим

$$z = \ln\left(\frac{1}{(n-1)!} z^{n-1} + \frac{1}{(n-2)!} z^{n-2} + \dots + \frac{1}{1!} z + 1\right) - \ln d^B.$$

Можно показать, что нелинейный оператор, стоящий в правой части этого уравнения – сжимающий. Следовательно, возможно построение итерационной процедуры

$$z_{k+1} = \ln\left(\frac{1}{(n-1)!} z_k^{n-1} + \frac{1}{(n-2)!} z_k^{n-2} + \dots + \frac{1}{1!} z_k + 1\right) - \ln d^B, \quad k=0,1,2,\dots$$

(с начальным значением $z_0 = 2n$). В пределе

$$z = \lim_{k \rightarrow \infty} z_k. \quad (5.39)$$

Именно таким способом были получены значения z , приведенные в табл. 5.2.

Таблица 5.2 - Значения корня z трансцендентного уравнения

$n \backslash d^B$	2	3	5	7	10	13	15
0,005	7,4301	9,2738	12,5941	15,6597	19,9984	24,1450	26,8360
0.01	6,6384	8,4059	11,6046	14,5706	18,7831	22,8208	25,4461
0,03	5,3559	6,9838	9,9610	12,7466	16,7312	20,5730	23,0800

Из выражения (5.34) легко получить выражение для максимальной загрузки B -соединения речевыми B -пакетами с учетом потерь b^{malty}

$$\widehat{\rho}_{st}^{Bp \max} = \left(1 - \frac{z}{\zeta}\right) = \left(1 - \frac{zL^{Bp}v^B}{v^B\theta_{st}^{Bp} - (L^{Bp} - H_{IP})} \frac{1}{V}\right). \quad (540)$$

Известно, что среднее время T_{ij}^C пребывания меченного требования из класса p в СМО типа $M/M/1$, с учетом введенных выше приоритетов для трафиков классов B и C дается выражением

$$T_{ij}^{Cp} = \frac{L^{Bp} + H_{IP}}{V} + \frac{\frac{L^{Cp} + H_{IP}}{V} \widehat{\rho}_{ij}^{Cp} + \frac{L^{Bp} + H_{IP}}{V} \widehat{\rho}_{ij}^{Bp}}{1 - \widehat{\rho}_{ij}^{Bp} - \widehat{\rho}_{ij}^{Cp}} + \widehat{\rho}_{ij}^{Bp} T_{ij}^{Cp}. \quad (541)$$

Общее среднее время пребывания ячейки данных T_{st}^{Cp} в n -звенном однородном канале передачи, включая задержку, связанную с накоплением информационной части ячейки в оконечной системе, дается выражением

$$T_{st}^{Cp} = n \frac{\frac{L^{Cp} + H_{IP}}{V}}{(1 - \widehat{\rho}_{st}^{Bp})(1 - \widehat{\rho}_{st}^{Cp})} + T_{pac}^{Cp}, \quad (542)$$

где $T_{pac}^{Cp} = \frac{L^{Cp} - H_{IP}}{\omega^C}$. Для однородного n -звенного канала $\widehat{\rho}_{ij}^{Cp} = \widehat{\rho}_{st}^{Cp}$ и дается выражением

$$\widehat{\rho}_{st}^{Cp} = 1 - \widehat{\rho}_{st}^{Bp} - \left[\frac{L^{Cp} + H_{IP}}{T_{st}^{Cp} - T_{pac}^{Cp}} \right] \frac{n}{(1 - \widehat{\rho}_{st}^{Bp})V}. \quad (543)$$

Отметим, что в отличие от низкоскоростных и среднескоростных ЦСИС в ИТС на технологии IP-QoS восстановление поврежденных сообщений данных осуществляется только на уровне ТСР.

Текущее значение $\widehat{\rho}_{ij,IP}^{Bp}$ можно вычислить следующим образом. Пропущенная суммарная изохронная нагрузка в канале $ij \in l_{st,m}^{Bp}$, полученная для заданной системы маршрутов R_{st}^{Bp} в сети, равна $\widehat{a}_{ij}^{Bp} = a_{ij}^{Bp} (1 - b_{ij}) = \sum_{st \in S^{B,q}} \sum_{m=1}^{M_{st}^{Bp}} p_{st,m}^{Bp} \widehat{a}_{st}^{Bp}$ (Эрл). Тогда эффективная скорость передачи, например, речевого В-трафика в канале $ij \in l_{st,m}^{Bp}$ определяется соответственно как

$$V_{ij,IP}^{Bp} = \widehat{a}_{ij}^{Bp} z^{Bp} \eta^{Bp} v^B, \quad (544)$$

Здесь z^{Bp} - коэффициент сжатия речевой информации; η^{Bp} - коэффициент уплотнения пауз в речевом соединении; Типичное значение для речи $\eta^{Bp} = 0,497$ [18]. С другой стороны требуемая минимальная эффективная скорость переноса речевого трафика в канале $ij \in l_{st,m}^{Bp}$ с учетом выводов раздела 2 вычисляется через общий критерий

$$\text{эффективности } K_{ij,IP}^{Bp} = \frac{L^{Bp} - H_{IP}}{L^{Bp}} \rho_{ij}^{Bp} :$$

$$V_{ij,IP}^{Bp} = V_{ij} K_{ij,IP}^{Bp} = V_{ij} \frac{(L^{Bp} - H_{IP}) \rho_{ij}^{Bp}}{L^{Bp}}. \quad (5.45)$$

Приравнявая (5.44) и (5.45) и разрешив уравнение относительно ρ_{ij}^{Bp} , получаем удельную загрузку системы речевым трафиком

$$\rho_{ij,IP}^{Bp} = \frac{L^{Bp}}{L^{Bp} - H_{IP}} \frac{v^{Bp}}{V_{ij}} \widehat{a}_{ij}^{Bp} z^{Bp} \eta^{Bp}. \quad (5.46)$$

Ясно, что должно выполняться неравенство

$$\frac{L^{Bp}}{L^{Bp} - H_{IP}} \frac{v^{Bp}}{V_{ij}} \widehat{a}_{ij}^{Bp} z^{Bp} \eta^{Bp} \leq \rho_{ij,IP}^{Bp \max}, \quad (5.47)$$

где $\rho_{ij,IP}^{Bp \max}$ – максимально возможная удельная загрузка ЛЦТ речевым трафиком. При сделанных предположениях и известных значениях $\rho_{ij,IP}^{Bp}$ и L^{Bp} можно получить выражение для нахождения значения $\rho_{ij,IP}^{Cp}$, пропущенного в рамках установленных мультимедийных соединений.

Известно [28], что среднее время пребывания пакета данных класса C в звене, моделируемом СМО типа $M/M/1$ с абсолютным приоритетом обслуживания речевого трафика, дается выражением:

$$T_{ij}^{Cp} = \frac{L^{Cp}(1 - \rho_{ij}^{Bp}) + L^{Bp} \rho_{ij}^{Bp}}{(1 - \rho_{ij}^{Bp} - \rho_{ij}^{Cp})(1 - \rho_{ij}^{Bp})V_{ij}}. \quad (5.48)$$

Решая это уравнение относительно ρ_{ij}^{Cp} , получаем

$$\rho_{ij,IP}^{Cp} = (1 - \rho_{ij}^{Bp} - \frac{L^{Cp}}{V_{ij} T_{ij}^{Cp}} + \frac{\rho_{ij}^{Bp}}{1 - \rho_{ij}^{Bp}} \frac{L^{Bp}}{V_{ij} T_{ij}^{Cp}}). \quad (5.49)$$

Аналитическое описание логических соединений межсетевого уровня для речевого трафика и трафика данных с учетом протокольной избыточности соответственно даются следующими выражениями

$$V_{ij,IP}^{Bp} = V \frac{L^{Bp} - H_{IP}}{L^{Bp}} \rho_{ij,IP}^{Bp \max} \quad (5.50)$$

и

$$V_{ij,IP}^{Cp} = V \frac{L^{Cp} - H_{IP}}{L^{Cp}} (1 - \rho_{ij}^{Bp} - \frac{L^{Cp}}{V_{ij} T_{ij}^{Cp}} + \frac{\rho_{ij}^{Bp}}{1 - \rho_{ij}^{Bp}} \frac{L^{Bp}}{V_{ij} T_{ij}^{Cp}}). \quad \dots\dots\dots (5.28)$$

5.5.3. Аналитическое описание логических соединений уровня сетевого доступа

Аналитическое описание логических соединений уровня сетевого доступа для речевого трафика и трафика данных с учетом протокольной избыточности, вносимой соответствующими технологиями (MPLS, FR и др.) могут быть представлены в следующем виде

$$V_{ij,NA}^{Bp} = V \frac{L^{Bp}}{L^{Bp} + H_{NA}}, \quad (5.29)$$

$$V_{ij,NA}^{Cp} = V \frac{L^{Cp}}{L^{Cp} + H_{NA}}. \quad (5.30)$$

5.6. Аналитические модели логических соединений ЛЦТ и составного тракта передачи IP-QoS-ИТС

Аналитические модели логических соединений для речевого трафика и трафика данных в ЛЦТ и составном тракте передачи в ИТС на технологии IP-QoS с учетом мультипликативной формы представления $K_{ij}^{k_p} = \prod_{h=1}^4 K_{h,ij}^{k_p}$, и результатов, полученных в разд. 5.3, имеют вид:

$$\begin{aligned} V_{ij}^{Bp} &= V \frac{L^{Bp} - H_{IP}}{L^{Bp} + H_{NA}} \rho_{ij}^{Bp}, \\ V_{ij}^{Cp} &= V \frac{L^{Cp} - H_{IP}}{L^{Cp} + H_{NA}} \left(1 - \rho_{ij}^{Bp} - \frac{L^{Cp}}{V_{ij} T_{ij}^{Cp}} + \frac{\rho_{ij}^{Bp}}{1 - \rho_{ij}^{Bp}} \frac{L^{Bp}}{V_{ij} T_{ij}^{Cp}} \right), \end{aligned} \quad (5.31)$$

$$\begin{aligned} V_{st}^{Bp} &= V \frac{L^{Bp} - H_{IP}}{L^{Bp} + H_{NA}} \sum_{m=1}^{M_{st}^{Bp}} p_{st,m}^{Bp} r_{st,m}^{Bp} \sqrt{\prod_{ij \in I_{st,m}^{Bp}} \frac{L^{Bp} \hat{a}_{ij}^{Bp} v^{Bp} \eta^B z^B}{L^{Bp} - H_{IP}}}, \\ V_{st}^{Cp} &= V \frac{L^{Cp} - H_{IP}}{L^{Cp} + H_{NA}} \sum_{m=1}^{M_{st}^{Cp}} p_{st,m}^{Cp} r_{st,m}^{Cp} \sqrt{\prod_{ij \in I_{st,m}^{Cp}} \left(1 - \rho_{ij}^{Bp} - \frac{L^{Cp}}{V_{ij} T_{ij}^{Cp}} + \frac{\rho_{ij}^{Bp}}{1 - \rho_{ij}^{Bp}} \frac{L^{Bp}}{V_{ij} T_{ij}^{Cp}} \right) \left(-\frac{p_0}{1 - p_0} \ln p_0 \right)}. \end{aligned} \quad (5.32)$$

Подчеркнем еще раз, что общие сетевые функционалы оценки эффективности использования ресурса пропускной способности системы асинхронным трафиком имеет ярко выраженную условную зависимость от изохронного трафика.

Коэффициент относительного использования цифрового тракта на транспортном уровне дается выражением с учетом квантильного подхода для оценки качества обслуживания мультимедийного соединения

$$R^p = \left[V_{ij}^{\min Bp} (1 - d^{Bp}) + V_{ij}^{\min Cp} (1 - d^{Cp}) \right] / V_{ij} \quad (5.33)$$

5.7. Аналитическое описание показателей качества IP-QoS-ИТС

Выше с учетом «квантильного» подхода были определены основные ограничения на качественные обслуживания мультимедийных соединений. При оптимизации функционала для речевого трафика в качестве ограничений будет выступать выражение (4.2). Если обозначить $T_{st,m}^{Bp}$ случайную задержку речевого пакета в m -м виртуальном

канале множества L_{st}^{Bp} , T_{st}^{Bp} – случайную величину вида $T_{st}^{Bp} = \sum_{m=1}^{M_{st}^{Bp}} p_{st,m}^{Bp} T_{st,m}^{Bp}$, а соответствующие им функции распределения $F_{st,m}^{Bp}(t) = \Pr\{T_{st,m}^{Bp} \leq t\}$ и

$F_{st}^{Bp}(t) = \Pr\{T_{st}^{Bp} \leq t\}$, то с учетом предполагаемой независимости
 $F_{st}^{Bp}(t) = \sum_{m=1}^{M_{st}^{Bp}} P_{st,m}^{Bp} F_{st,m}^{Bp}(t)$. Соответствующая плотность распределения
 $f_{st}^{Bp}(t) = \sum_{m=1}^{M_{st}^{Bp}} P_{st,m}^{Bp} f_{st,m}^{Bp}(t)$, где $f_{st,m}^{Bp}(t) = (F_{st,m}^{Bp}(t))'$. Искомая вероятность превышения
задержки $\Pr_{st}(t \geq \theta_{st}^{Bp}) = \sum_{m=1}^{M_{st}^{Bp}} P_{st,m}^{Bp} \int_{\theta_{st}^{Bp}}^{\infty} f_{st,m}^{Bp}(t) dt$. Интегрируя и подставляя пределы
интегрирования, получаем

$$\Pr_{st}(t \geq \theta_{st}^{Bp}) = \sum_{m=1}^{M_{st}^{Bp}} P_{st,m}^{Bp} (1 - F_{st,m}^{Bp}(\theta_{st}^{Bp})), \quad \forall st \in S^B : \hat{a}_{st}^{Bp} \neq 0. \quad (5.34)$$

Квантили $F_{st,m}^{Bp}(\theta_{st}^{Bp})$ легко вычисляются при принятых предположениях.

При оптимизации функционала для трафика данных с учетом, что параметры трафика речи оптимальны и фиксированы, в качестве основного ограничения (4.4) выступает ограничение на заданное среднее время доставки в сеансе связи T_{st}^{Cp} пакетов данных класса C (или (4.5) при «квантильном» подходе – на $\sum_{m=1}^{M_{st}^{Cp}} \rho_{st,m}^{Cp} (1 - F_{st,m}^{Cp}(T^{Cp})) \leq d^{Cp}$). С учетом выше сформулированных допущений это ограничение принимает следующий вид:

$$T_{st}^{Cp} = \sum_{m=1}^{M_{st}^{Cp}} P_{st,m}^{Cp} T_{st,m}^{Cp} - \frac{L^{Cp} - H_{IP}}{\omega^C} = \sum_{m=1}^{M_{st}^{Cp}} P_{st,m}^{Cp} \left(\sum_{ij \in I_{st,m}^C} T_{ij}^{Cp} + \sum_{\forall i: ij \in I_{st,m}^C} T_i^{Cp} \right) - \frac{L^{Cp} - H_{IP}}{\omega^C} \leq T^{Cp}. \quad (5.35)$$

Для всех $st \in S^{Cp} : \hat{a}_{st}^{Cp} \neq 0$ – это основное ограничение задачи анализа пакетной ИТС на технологии IP-QoS для трафика данных класса C .

Полученные выше результаты позволяют сформулировать задачу анализа инфокоммуникационной сети на технологии IP в общем виде для неоднородных сетей следующим образом.

1. При заданной топологии сети, структуре потоков, заданной в виде матриц тяготений Y^k и заданной системе маршрутов найти значения $\rho_{ij}^{Bp \max}$ и $L^{B \max}$, доставляющих максимум функционалу

$$\arg \max K_{st}^{Bp}, \text{ при условиях}$$

$$\sum_{m=1}^{M_{st}^{Bp}} P_{st,m}^{Bp} (1 - F_{st,m}^{Bp}(\theta_{st}^{Bp})),$$

$$L^{Bp \max} \triangleleft \theta_{st}^B v^B - H_{IP}, 0 \triangleleft \rho_{ij}^{Bp \max} \leq 1 \quad \forall st \in S^{Bp} : \hat{a}_{st}^{Bp} \neq 0 \quad (5.35)$$

2. При найденных значениях $\rho_{ij}^{Bp \max}$ и $L^{B \max}$ найти значения $\rho_{ij}^{Cp \max}$ и $L^{C \max}$, доставляющих максимум функционалу

$$\arg \max K_{st}^C,$$

при ограничениях $T_{st}^{Cp} = \sum_{m=1}^{M_{st}^{Cp}} P_{st,m}^{Cp} T_{st,m}^{Cp} - \frac{L^{C \max} - H_{IP}}{\omega^C} \leq T^{Cp} \quad \forall st \in S^{Cp} : \hat{a}_{st}^{Cp} \neq 0$ или

$$\sum_{m=1}^{M^{Cp}} \rho_{st,m}^{Cp} (1 - F_{st,m}^{Cp}(T^{Cp})) \leq d^{Cp} \text{ (при «квантильном» подходе).}$$

Пара $(\hat{a}_{ij}^{Bp}, V_{ij}^{Cp})$ характеризует эффективность передачи смешанного трафика по тракту st инфокоммуникационной сети на технологии IP с заданным качеством обслуживания.

Замечание 1. Результаты задачи анализа ИТС лежат в основе процедуры синтеза сети при поиске минимальных значений V_{ij}^k , при условии пропускания заданной интегральной нагрузки с требуемым качеством передачи и минимальной стоимости каналов связи.

Замечание 2. Учитывая, что весовой коэффициент виртуального соединения на транспортном уровне определяется долей входящего потока \hat{a}_{st}^k в общем входящем потоке сети $\hat{Y}^k = \sum_{st \in S^k} \hat{a}_{st}^k$, можно ввести понятие средневзвешенного по потокам общего функционала использования пропускной способности инфокоммуникационной сети произвольной топологии разнородным трафиком, который имеет вид

$$K^k = q^k \sqrt{\sum_{st \in S^k} \frac{\hat{a}_{st}^k}{\hat{Y}^k} (K_{st}^k)^2}, \quad q^k = |S^k|. \text{ Функционал подобного типа обладает свойством}$$

$$\lim_{q \rightarrow \infty} K^k = \max_{st \in S^k} K_{st}^k, \quad q^k = |S^k|, \text{ что не дает явного преимущества транспортным TCP-}$$

соединениям с большими весовыми коэффициентами. Тогда указанные задачи могут решаться также одновременно и для всей сети с использованием средневзвешенных по потокам функционалов использования сети K^{Bp} и K^{Cp} , которые должны максимизироваться при условиях того же вида, но уже для всех $st \in S^k$, имеющих ненулевые потоки.

Вопросы для самопроверки по теме 5.

1. Обоснуйте качественные показатели МСС
2. Обоснуйте функциональный критерий эффективности МСС.
3. Системная модель инфотелекоммуникационной транспортной системы
4. Поясните принцип функционально-структурной целостности инфотелекоммуникационных транспортных систем
5. Концепция архитектуры мультисервисной сети и ее основные особенности
6. Опишите логическую структуру архитектуры инфотелекоммуникационных транспортных систем на технологи IP-QoS
7. Опишите логическую структуру архитектуры инфотелекоммуникационной транспортной ATM-системы. Реализация базовых сетевых функций и основных протоколов в модели архитектуры BPRM
8. Поясните принцип уровневой иерархии МСС. Укажите общий подход к построению критерия эффективности на базе архитектуры МСС
9. Опишите процесс формализации критериев эффективности ИТС-IP-QoS
10. Опишите процесс формализации критериев эффективности ИТС-ATM
11. Поясните принцип единственности МСС как основу задачи ее анализа
12. Сформулируйте задачу анализа ИТС в общем виде

