

Передача видео по беспроводным сетям¹

Антон Кирьянов, Евгений Хоров (ИППИ РАН)

В последние годы во всем мире отмечается колоссальный рост объемов передаваемой мультимедийной информации, связанный с развитием Интернета и ростом рынка мультимедийных услуг. Согласно прогнозам ведущих телекоммуникационных компаний Cisco и AT&T к 2015 году доля мультимедийных данных в пользовательском трафике превысит 60%, причем большая его часть будет представлять собой видео высокого разрешения, передаваемое в режиме реального времени. Помимо требований к широкой полосе, такой тип трафика накладывает ограничения на время доставки пакета и вероятность его потери, то есть предъявляет определенные требования к качеству обслуживания (Quality of Service) – QoS-требования.

Для обеспечения доступа абонентов к широкополосным мультимедийным услугам сегодня все более и более популярным становится использование беспроводных сетей. Существенным недостатком беспроводного канала является наличие случайных помех и замирания сигнала, приводящих к тому, что канал оказывается нестационарным, т.е. вероятность успешной попытки передачи пакета данных, время доставки пакетов и пропускная способность канала меняются со временем. Это затрудняет выполнение требований к качеству обслуживания при предоставлении доступа к широкополосным мультимедийным услугам, в том числе и при передаче мультимедийных данных реального времени, например, аудио- и видеопотоков. Требования к обслуживанию таких потоков в общем случае заключаются в обеспечении низкого времени доставки пакетов, а также низкой доли не доставленных пакетов.

Традиционно для записи, хранения, передачи и воспроизведения динамического изображения оно представляется в виде потока (последовательности) видеок кадров (статических изображений), как правило, полученных с некоторой частотой дискретизации. Эта частота в общем случае может быть переменной, однако на практике она чаще всего составляет около 25-30 кадров в секунду, что обусловлено физиологическими особенностями человеческого восприятия.

Помимо собственно последовательности видеок кадров, называемой далее видеоизображением, мультимедийный поток может включать в себя один или несколько аудиопотоков, субтитры и другую информацию.

Одной из наиболее часто используемых технологий представления мультимедийного потока в цифровом виде является технология MPEG-4. Несколько десятков документов, соответствующих разным частям стандарта ISO/IEC 14496, описывают различные аспекты этой технологии, в том числе, структуру мультимедийного потока, аудио и видео кодеки, работу с субтитрами, алгоритмы сжатия информации, контейнеры, методы тестирования на совместимость частей стандарта и т.д.

Технология MPEG-4 использует иерархическую структуру с вложенными друг в друга сущностями. Сущностью самого верхнего уровня является последовательность визуальных

¹ Материал подготовлен для ознакомления студентов ФРТК с особенностями кодирования видеоизображения и проблемами, возникающими при передаче видео по беспроводным сетям

объектов (Visual Object Sequence). Внутри визуального объекта (Visual Object), как правило, содержится слой видеоизображения.

Слой видеоизображения представляет собой сущность, содержащую группу (Group of Pictures, GoP) предметных плоскостей видеоизображения (Video Object Plane, VOP), каждая из которых соответствует одному видеокадру.

Как правило, два идущих подряд видеокадра незначительно отличаются друг от друга, поэтому для сжатия информации оказывается эффективным вместо кодирования последующего видеокадра целиком, закодировать только изменения этого видеокадра относительно предыдущего.

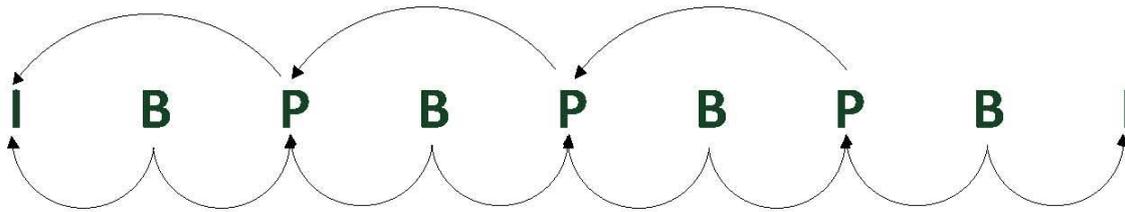
MPEG-4 Part2 использует 3 основных типа предметных плоскостей:

- I-VOP (Intra-coded) – предметная плоскость, содержащая полную информацию о видеокadre; может быть декодирована корректно независимо от остальных кадров. Именно с I-VOP начинается новая группа предметных плоскостей.
- P-VOP (Predictive-coded) – предметная плоскость, содержащая не полную информацию о видеокadre, а только изменения относительно прошлой (past reference) P- или I-VOP, называемой базовой VOP; может быть декодирован корректно, только если базовая VOP декодирована корректно.
- B-VOP (Bidirectionally predictive-coded) – предметная плоскость, содержащая не полную информацию о видеокadre, а только изменения относительно прошлой и будущей базовой VOP (P- или I-VOP); может быть декодирована корректно, только если корректно декодированы обе (прошлая и будущая) базовые VOP.

Разный объем информации, в различных предметных плоскостях приводит к тому, что P-VOP имеет, как правило, меньший размер, чем I-VOP, а B-VOP имеет меньший размер, чем P-VOP. Из этого следует, что для повышения степени сжатия информации при создании мультимедийного потока MPEG-4, кодеку следует генерировать I-VOP как можно реже. На практике интервал между двумя последовательными I-VOP ограничивают сверху, чтобы снизить объем данных, которые надо получить и обработать, для того чтобы корректно отобразить видеоизображение. Кроме того, кодеки автоматически генерируют I-VOP при изменении сцены (плана). Это приводит к тому, что длина группы предметных плоскостей, чаще всего оказывается случайной величиной.

Помимо длины группы предметных плоскостей, случайной может быть и структура группы предметных плоскостей (GoP), которая может подбираться автоматически исходя из характера видеоизображения.

На рисунке ниже показана GoP размером в 7 кадров, включающая в себя одну I-VOP, три P-VOP и три B-VOP. Следующая за ними I-VOP относится к следующей группе предметных плоскостей.



При передаче мультимедийного потока по сети его необходимо упаковать (инкапсулировать) в кадры транспортного уровня. Сложная структура мультимедийного потока и, в общем случае, ненадежная передача в сети делают задачу инкапсуляции мультимедийного потока в кадры транспортного уровня нетривиальной.

Большая востребованность передачи мультимедийных данных и интерес к этой проблеме со стороны различных сообществ исследователей и разработчиков привели к появлению принципиально разных протоколов передачи мультимедийных потоков, значительная часть из которых используется и сегодня. Выбор протоколов обусловлен требованиями со стороны оказываемого сервиса.

Применительно к передаче видеоизображения, выделяют следующие основные сервисы.

1. «Воспроизведение видеофрагмента, размещенного на сервере». Данный сервис означает возможность пользователям открыть и проиграть размещенный на сервере видеофрагмент без предварительного копирования его на машину клиента. Этот видеофрагмент может быть размещен на файловом сервере или веб-сервере.
2. «Воспроизведение видеоизображения в режиме реального времени». Данный сервис используется в таких приложениях как Live television (телевидение в режиме реального времени), видеоконференции и др. Существенным отличием этого сервиса от предыдущего является то, что видеоизображение не существует в виде заранее созданного видеофрагмента, а генерируется на сервере в режиме реального времени. Это накладывает жесткие ограничения на время доставки пакетов видеопотока, что не позволяет использовать буфер большого размера. Как следствие, страдает качество видеоизображения: в случае невозможности доставить пакеты видеопотока они отбрасываются.

Для предоставления сервиса «Воспроизведение видеофрагмента, размещенного на сервере», как правило, используются технологии, гарантирующие доставку мультимедийного потока без потерь.

С точки зрения пользователя проигрывание видеофрагмента, размещенного на удаленном сервере, например, на сервере samba, слабо отличается от проигрывания видеофрагмента, находящегося на жестком диске самого клиента. Единственным отличием является то, что при чтении очередной порции данных проигрывателем происходит обращение не к локальному жесткому диску, а к удаленному серверу. Для этого используются специально разработанные протоколы передачи файлов по сети, например, протокол Server Message Block (SMB), который является протоколом уровня приложения и работает поверх протоколов транспортного уровня, гарантирующих доставку пакетов, переданных по сети. В качестве протокола транспортного уровня, как правило, используется протокол Transmission Control Protocol (TCP).

Развитие Интернета дало толчок для внедрения технологии HTTP Streaming, которая представляет собой передачу видеоизображения с использованием протокола HTTP. HTTP

(HyperText Transfer Protocol, «протокол передачи гипертекста») — протокол прикладного уровня передачи данных (изначально — в виде гипертекстовых документов). Основой HTTP является технология «клиент-сервер», то есть предполагается существование клиентов, которые инициируют соединение и посылают запрос, и серверов, которые ожидают соединения для получения запроса, производят необходимые действия и возвращают обратно сообщение с результатом.

Одной из реализаций HTTP Streaming является технология HTTP progressive download, которая представляет собой загрузку видеоизображения в проигрыватель по протоколу HTTP и, по сути, является обычной загрузкой данных с сервера. После начала загрузки многие приложения отображают полосу загрузки, маркер текущей позиции, и как только будет загружена достаточная для декодирования порция данных, начинают воспроизведение видеоизображения. При этом если канал достаточно узкий, через некоторое время проигрывание видеоизображения будет остановлено до тех пор, пока не загрузится очередная, достаточная для декодирования порция данных.

При использовании HTTP progressive download видеоизображение будет в конечном итоге загружено в память устройства именно в том виде, в котором оно хранится на сервере и не будет содержать в себе никаких артефактов, что несомненно является плюсом данной технологии. В то же время загрузка видеоизображения может занять длительное время, и при узком канале просматривать видеоизображение в режиме реального времени не удастся.

Несмотря на значительно большую сложность технологий адаптивной передачи с использованием HTTP по сравнению с традиционными способами передачи видеоизображения, для решения ряда задач технологии адаптивной передачи на практике оказываются более привлекательными из-за масштабируемости, широкой распространенности протокола HTTP и отсутствия необходимости развертывания специальных серверов, которые требуются для осуществления традиционного потокового вещания.

Технология HTTP Streaming широко используется в сценарии «Воспроизведение видеофрагмента, размещенного на сервере», так как предоставляет гарантию целостной передачи всего видеоизображения, но для этого может потребоваться длительное время. Из-за этого HTTP Streaming нецелесообразно использовать в сценарии «Воспроизведение видеоизображения в режиме реального времени», где время доставки видеоизображения до конечного получателя существенно ограничено.

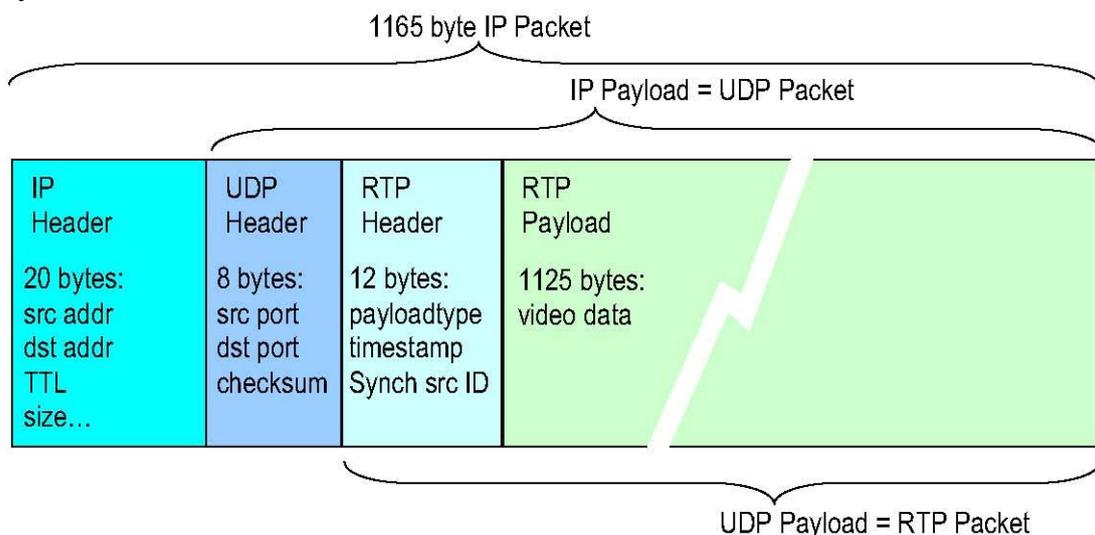
Для этих задач были разработаны другие технологии передачи, примером которых являются технологии, не гарантирующие доставку мультимедийного потока без потерь. Для этого в качестве протокола транспортного уровня, как правило, используется протокол UDP, который не гарантирует доставку пакета до конечного получателя. Ценность данного протокола заключается в его простоте и скорости работы – так как протокол не осуществляет гарантию доставки пакетов, то не тратится время и ресурсы сети на получение подтверждения о доставке пакета, а также на повторные передачи, если предыдущая попытка оказалась неудачной. Поэтому чувствительные к временным задержкам сервисы часто используют в качестве протокола транспортного уровня протокол UDP. Так, например, при передаче видеоизображения в режиме реального времени целесообразнее попробовать доставить новый видеокادر, а не совершать очередные попытки передачи уже «устаревших» видеокладов.

Технология Raw UDP базируется на упаковке данных кодера напрямую в пакеты протокола UDP. К сожалению, у такой технологии несмотря на ее простоту имеется большое число

недостатков. Протокол UDP не гарантирует правильный порядок пакетов, который необходим для правильного декодирования видеоизображения, а также не содержит отметки времени, которые позволяют синхронизировать относительные временные сдвиги времени между источником и получателем пакетов. В силу отсутствия механизмов обратной связи, использование одного лишь протокола UDP не позволяет получать контроль за качеством обслуживания, не позволяет обнаруживать и устранять возникающие в сети проблемы для улучшения качества обслуживания конечных пользователей.

Из-за перечисленных выше причин технология Raw UDP не рекомендуется к использованию при передаче видеоизображения.

Для устранения недостатков, возникающих при использовании технологии Raw UDP, была разработана технология RTP + UDP. Технология RTP + UDP представляет собой прямую упаковку данных кодека MPEG-4 Part2 в пакеты RTP.



(Изображение с сайта <http://www.envivio.com/>)

Протокол сеансового уровня RTP, разработанный одной из рабочих групп Инженерного совета Интернета (IETF), широко используется при передаче мультимедийных данных и, как правило, функционирует совместно с протоколом транспортного уровня UDP.

Использование протокола RTP открывает многие возможности, которые оказываются чрезвычайно полезными при передаче видеопотоков по сети.

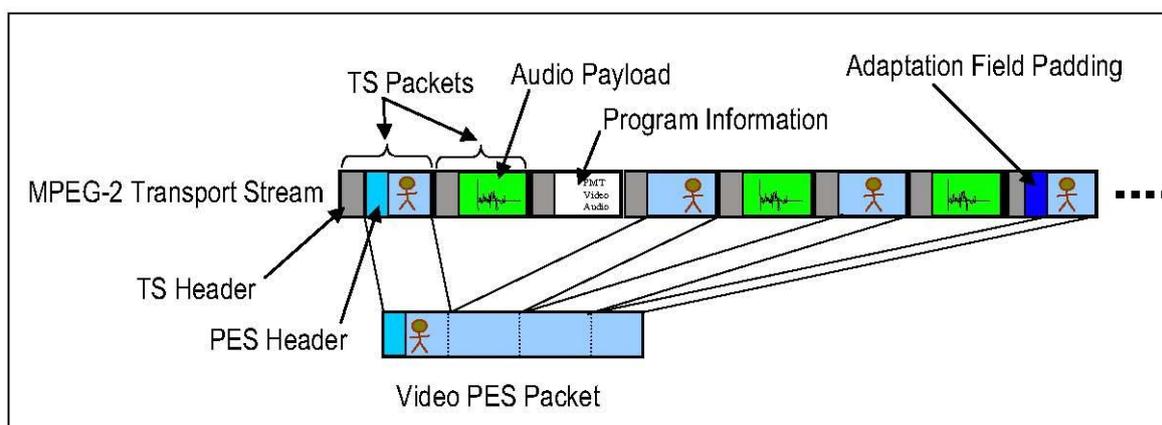
Инкапсуляция RTP + UDP является достаточно прозрачной для определения структуры MPEG потока, включая границы между видеокадрами и их типы. Даже если I-VOP или какие-то другие кадры не маркируются приложением, то определить их тип позволяют уникальные шестнадцатеричные последовательности, которые располагаются вначале каждого видеокadra и позволяют однозначно определить тип данного видеокadra. Даже если видеокadra оказывается фрагментирован на несколько UDP пакетов, то благодаря нумерации RTP пакетов не составляет труда определить, часть видеокadra какого типа передается в данном UDP пакете.

Для передаче мультимедийного потока по сети перед упаковкой в кадры транспортного уровня часто применяется технология, формирующая специализированный транспортный поток MPEG Transport Stream (MPEG-TS).

Поток MPEG Transport Stream представляет собой мультиплексированные данные нескольких потоков видео, аудио, субтитров, служебной информации.

Формат разрабатывался как универсальный способ мультиплексирования MPEG данных для передачи их по сети, который будет эффективно работать с большинством протоколов транспортного, сетевого и канального уровней. В MPEG Transport Stream предусмотрены механизмы для синхронизации потоков, а также наличие избыточной информации, которая позволяет исправлять возникающие при передаче по ненадежному соединению ошибки методом упреждения.

Согласно спецификации MPEG-TS видеопотоки, аудиопотоки, субтитры, а также потоки служебной информацией, каждый из которых носит название элементарного потока (Elementary Stream - ES), формируют единый транспортный поток MPEG-TS. Каждый элементарный поток затем на уровне приложения разбивается на пакеты, превращаясь в пакетированный элементарный поток (Packetized elementary stream). Например, единица пакетированного элементарного потока (PES пакет) с видеоданными содержит в себе один видеокادر. PES пакеты могут иметь разную длину, поэтому они разбиваются на так называемые пакеты транспортного потока (TS пакеты) фиксированной длины.



(Изображение с сайта <http://www.envivio.com/>)

Отметим, что MPEG TS представляет собой передачу данных без обратной связи, то есть не содержит механизмов оповещения источника о доле потерянных пакетов, джиттере, и другой статистике, что позволило бы адаптировать передачу к текущим условиям и улучшить качество передачи потока. С другой стороны, так как MPEG TS – однонаправленный поток, то не требуется обмена информацией между источником и приемником, в том числе и установки логического соединения между ними, в результате чего процесс передачи сводится к простому вещанию данных от сервера.

К сожалению, сам принцип формирования потока MPEG TS порождает и трудности. При формировании единого MPEG Transport Stream потока TS пакеты от разных потоков (видео, аудио, субтитры) перемешиваются между собой, причем один PES пакет, содержащий видеокادر, может разбиваться на несколько TS пакетов, которые впоследствии могут быть упакованы в различные пакеты на транспортном уровне. По той же причине в один пакет транспортного уровня могут быть упакованы части различных PES пакетов. Все это приводит к тому, что в один пакет канального уровня попадают пакеты разных PES пакетов, а в разные пакеты канального уровня могут попадать части одного и того же PES пакета. Таким образом, при потере одного пакета на канальном уровне может происходить потеря сразу нескольких PES пакетов, в результате чего теряется намного больший объем полезных данных, чем размер одного пакета на канальном уровне. Более того, в таком виде становится практически невозможным произвести эффективную приоритезацию пакетов на канальном уровне для повышения качества передачи видеоизображения.

Несмотря на это, MPEG-TS широко используется в индустрии на всех этапах деятельности, начиная от массового вещания и заканчивая пользовательскими приставками. Причина этого заключается в том, что мультиплексирование и демупльтиплексирование потоков может быть сделано с использованием оборудования, а не программного обеспечения, а также в отсутствии необходимости использования IP сетей, что означает простоту применения такого подхода при использовании спутникового телевидения, DVD проигрывателей и других различных пользовательских устройств.

Несмотря на большое разнообразие методов передачи видеопотоков, в настоящее время все еще остается открытыми множество вопросов, связанных с надежной доставкой видеопотоков реального времени по беспроводным сетям.

Как уже было отмечено в начале статьи, для беспроводных сетей характерны случайные помехи, приводящие к тому, что качество канала меняется со временем. Это делает необходимой разработку методов повышения надежности передачи видеопотоков, что можно сделать разными способами. В настоящее время в мире проводятся исследования по двум направлениям: 1) повышение надежности передачи отдельных кадров (независимо от их места в структуре видеопотока) и 2) фильтрация пакетов, с помощью которой важные пакеты передаются с большим приоритетом, чем менее важные.