

Е.И. ПРОХОРЕНКО

О ПАКЕТНОЙ ПЕРЕДАЧЕ РЕЧИ

Розглянуті існуючі можливості реалізації протоколів пакетної передачі мови в середовищі Internet для підвищення ефективності застосування пропускної спроможності IP-каналу. Найбільш перспективні продукти для IP-телефонії задовольняють рекомендаціям сімейства "G" стандарту H.323.

In article are considered existing possibility of realization of protocols packet issue a speech in ambience Internet for increasing efficiency of use reception capacity IP-channel. In the same way review of modern products is made for IP-telephony a family "G" standard N.323 described by recommendations.

Постановка проблемы. В последние годы в мире наблюдается стремительный рост объема передаваемых данных, что привело к бурному развитию высокоскоростных транспортных сетей, строящихся с применением IP-технологий [1 – 4]. Отчасти это обусловлено популярностью сети Интернет и очевидной выгодой построения универсальной сети транспортировки не только телефонного трафика, но и данных произвольного вида и структуры. Организация передачи речевого трафика непосредственно через сеть IP сопряжена с множеством проблем, которые вытекают из естественных особенностей телефонного разговора. Основное нежелательное явление – задержка передачи речевого сигнала от одного абонента другому. Задержка вызывает два нежелательных явления – эхо и наложение речи. Общая задержка в сети является величиной, состоящей, в том числе, из следующих компонентов: а) задержка накопления, которая вызвана необходимостью подготовки кадра из последовательности речевых отсчетов обрабатываемых вокодером; б) задержка кодирования, которая образуется непосредственно в процессе кодирования кадра цифровым процессором обработки сигналов (ЦПОС). Производительность ЦПОС должна быть такой, чтобы задержка кодирования была меньше или, по крайней мере, равна задержке накопления. Выбор ЦПОС можно сделать на основании данных о сложности применяемого алгоритма кодирования. Таким образом, на сегодняшний день существует проблема выбора протокола, который, наряду с передачей традиционного трафика сети передачи данных, обеспечивал бы качественную передачу речевого сигнала.

Цель статьи. Рассмотреть существующие возможности реализации протоколов пакетной передачи речи в среде Интернет для повышения эффективности использования пропускной способности IP-канала.

Методы кодирования речевой информации. Одним из важных факторов эффективного использования пропускной способности IP-канала является выбор оптимального алгоритма кодирования/декодирования речевой информации – кодека.

Все существующие сегодня типы речевых кодеков по принципу действия можно разделить на три группы:

- Кодеки с импульсно-кодовой модуляцией (ИКМ) и адаптивной дифференциальной импульсно-кодовой модуляцией (АДИКМ), использующиеся сегодня в системах традиционной телефонии. В большинстве случаев, представляют собой сочетание АЦП/ЦАП.
- Кодеки с вокодерным преобразованием речевого сигнала возникли в системах мобильной связи для снижения требований к пропускной способности радиотракта. Эта группа кодеков использует гармонический синтез сигнала на основании информации о его вокальных составляющих – фонемах. В большинстве случаев, такие кодеки реализованы как аналоговые устройства.
- Комбинированные (гибридные) кодеки сочетают в себе технологию вокодерного преобразования/синтеза речи, но оперируют уже с цифровым сигналом посредством специализированных DSP. Кодеки этого типа содержат в себе ИКМ или АДИКМ кодек и реализованный цифровым способом вокодер.

В голосовых шлюзах IP-телефонии понятие кодека подразумевает не только алгоритмы кодирования/декодирования, но и их аппаратную реализацию. Большинство кодеков, используемых в IP-телефонии, описаны рекомендациями семейства «G» стандарта H.323 [4, 5].

Преобразование аналогового речевого сигнала в цифровой вид обычно осуществляется методом импульсно-кодовой модуляции (ИКМ). После такой обработки речевой сигнал уже пригоден для передачи по цифровым каналам. Однако для передачи такого цифрового потока необходимо выделение полосы пропускания 64 Кбит/с (рекомендация ITU G.711 [6]), что является явно избыточным.

Рекомендация ITU-T G.711 [6]. Рекомендация, утверждённая МККТТ в 1984 г., описывает кодек, использующий ИКМ преобразование аналогового сигнала с точностью 8 бит, тактовой частотой 8 КГц и простейшей компрессией амплитуды сигнала. Скорость потока данных на выходе преобразователя составляет 64 Кбит/с. Для снижения шума квантования и улучшения преобразования сигналов с небольшой амплитудой, при кодировании используется нелинейное квантование по уровню. Кодек G.711 широко распространён в системах традиционной телефонии с коммутацией каналов. Несмотря на то, что рекомендация G.711 в стандарте H.323 является

основной и первичной, в шлюзах IP-телефонии данный кодек применяется редко из-за высоких требований к полосе пропускания и задержкам в канале передачи (всё-таки 64 Кбит/с – это много). Использование G.711 в системах IP-телефонии обосновано лишь в тех случаях, когда требуется обеспечить максимальное качество кодирования речевой информации при небольшом числе одновременных разговоров. Одним из примеров применения кодека G.711 могут послужить IP-телефоны компании CISCO.

Существует еще одно преобразование речевого сигнала – адаптивная дифференциальная импульсно-кодовая модуляция (ADPCM – Adaptive Differential Pulse Code Modulation). Этот алгоритм дает практически такое же качество воспроизведения речи, как и ИКМ, однако для передачи информации при его использовании требуется полоса всего в 16-32 Кбит/с (стандарт ITU G.726). Метод основан на том, что в аналоговом сигнале, передающем речь, невозможны резкие скачки интенсивности. Поэтому если кодировать не саму амплитуду сигнала, а ее изменение по сравнению с предыдущим значением, то можно обойтись меньшим числом разрядов. В ADPCM изменение уровня сигнала кодируется четырехразрядным числом, при этом частота измерения амплитуды сигнала сохраняется неизменной.

Рекомендация ИТУ-T G.726. Стандарт G.726 был принят в 1984 году. Метод базируется на цифровом кодировании сигнала, который получается путем преобразования и кодирования аналогового сигнала при помощи РСМ (Pulse Code Modulation). Трансформация аналогового речевого сигнала в цифровой РСМ-сигнал базируется на теореме Найквиста, в соответствии с которой цифровой сигнал, полученный выборкой из аналогового сигнала с частотой в два раза выше максимальной частоты, с помощью интерполяции можно обратно преобразовать в аналоговую форму. Человеческая речь воспроизводится с приемлемым качеством в полосе частот 200 – 3400 Гц, чему соответствует частота выборки 8 кГц, или 8000 отсчетов в секунду, а каждый отсчет преобразуется в 8-битовый цифровой код. Следовательно, общая скорость цифрового потока РСМ-сигнала равна 8×8000 отсчетов в секунду, что соответствует 64 Кбит/с. Хотя РСМ-сигнал со скоростью 64 Кбит/с и гарантирует качество речи аналогового телефонного сигнала, ограниченная общая ширина канала, особенно в спутниковых и радиочастотных системах, вынуждает снижать скорость битовых потоков, отводимых для каждого речевого сигнала. Скорость входного битового потока 64 Кбит/с ($8 \text{ кГц} \times 8\text{-битное РСМ слово}$). После ADPCM-кодирования скорости выходного потока могут быть 40, 32, 24 или 16 Кбит/с ($8 \text{ кГц} \times 5,4,3$ или 2-битное ADPCM слово). Кодек может применяться совместно с кодеком G.711 для снижения скорости кодирования последнего. Кодек предназначен для использования в системах видеоконференций.

Все методы кодирования, основанные на определенных предположениях о форме сигнала, эффективны лишь при скорости передачи цифрового

речевого сигнала большей, чем 32 Кбит/с. Дальнейшее снижение скорости возможно путем устранения избыточности, содержащейся в речевом сигнале, на основе анализа речи при передаче и воссоздания (синтеза речи) при приеме. В результате по каналу связи передают не собственно речь в цифровом виде, а ее характеристики, позволяющие восстановить речевой сигнал при приеме. Для выполнения процессов анализа-синтеза речи используют специальные кодеры-декодеры речевого сигнала, называемые вокодерами (сокращение от английских слов voice coder).

Наиболее эффективными являются вокодеры на основе метода линейного предсказания речи LPC (Linear Predictive Coding). Процесс кодирования в этом случае разбивается на ряд типовых процедур, каждую из которых выполняют методами цифровой фильтрации. При кодировании выявляют периодические процессы в речевом сигнале, определяют их параметры, а затем устраняют из речевого сигнала избыточность, исключая найденные периодичности.

В качестве входного сигнала в LPC используется последовательность цифровых значений амплитуды, но алгоритм кодирования применяется не к отдельным цифровым значениям, а к определенным их блокам. Для каждого такого блока значений вычисляются его характерные параметры: частота, амплитуда и ряд других. Именно эти значения и передаются по сети. При таком подходе к кодированию речи, во-первых, возрастают требования к вычислительным мощностям специализированных процессоров, используемых для обработки сигнала, а во-вторых, увеличивается задержка при передаче, поскольку кодирование применяется не к отдельным значениям, а к некоторому их набору, который перед началом преобразования следует накопить в определенном буфере.

Важно, что задержка в передаче речи связана не только с необходимостью обработки цифрового сигнала (эту задержку можно уменьшать, увеличивая мощность процессора), но и непосредственно с характером метода сжатия. Метод кодирования с линейным предсказанием LPC позволяет достигать очень больших степеней сжатия, которым соответствует полоса пропускания 2,4 или 4,8 Кбит/с, однако качество звука здесь сильно страдает. Поэтому в коммерческих приложениях он не используется, а применяется в основном для ведения служебных переговоров.

Рекомендация ITU-T G.729. G.729 – это CS-ACELP вокодер (Conjugate-structure Algebraic-code-excited Linear-Prediction). Алгоритм основан на модели кодирования с использованием линейного предсказания. Кодер оперирует с кадрами речевого сигнала длиной 10 мс, дискретизованными с частотой 8 КГц, что соответствует 80 16-битным отсчетам в линейном законе. Для каждого кадра производится анализ речевого сигнала и выделяются параметры: коэффициенты фильтра линейного предсказания, индексы и коэффициенты усиления в адаптивной и фиксированной кодовых книгах.

Далее это параметры кодируются в битовый поток и передаются в канал. В декодере эта битовая посылка используется для восстановления параметров сигнала возбуждения и коэффициентов синтезирующего фильтра. Далее восстанавливается речь, путем пропускания сигнала возбуждения через кратковременный синтезирующий фильтр. Вокодер обрабатывает кадры речевых сигналов длиной 10 мс. Дополнительно, существует задержка длиной 5 мс (look-ahead buffer), что в сумме выливается в 15 мс алгоритмическую задержку.

Более сложные методы сжатия речи основаны на применении LPC в сочетании с элементами кодирования формы сигнала. В этих алгоритмах используется кодирование с обратной связью, когда при передаче сигнала осуществляется оптимизация кода. Закодировав сигнал, процессор пытается восстановить его форму и сличает результат с исходным сигналом, после чего начинает варьировать параметры кодировки, добываясь наилучшего совпадения. Достигнув такого совпадения, аппаратура передает полученный код по линиям связи; на противоположном конце происходит восстановление звукового сигнала. Ясно, что для использования такого метода требуются еще более серьезные вычислительные мощности.

Одной из самых распространенных разновидностей описанного метода кодирования является метод LD-CELP (Low-Delay Code-Excited Linear Prediction). Он позволяет достичь удовлетворительного качества воспроизведения при пропускной способности 16 Кбит/с. Алгоритм применяется к последовательности цифр, получаемых в результате аналого-цифрового преобразования голосового сигнала с 16-разрядным разрешением. Пять последовательных цифровых значений кодируются одним 10-битовым блоком – это и дает те самые 16 Кбит/с. Для применения этого метода требуются большие вычислительные мощности; в частности, в марте 1995 г. ITU принял новый стандарт – G.723, который предполагается использовать при сжатии речи для организации видеоконференций по телефонным сетям. Этот стандарт представляет собой часть более общего стандарта H.324, описывающего подход к организации таких видеоконференций. Цель – организация видеоконференций с использованием обычных модемов.

Рекомендация ITU-T G.723.1 [7]. Описывает гибридные кодеки, использующие технологию кодирования речевой информации, сокращённо называемую – MP-MLQ (Multy-Pulse – Multy Level Quantization – Множественная Импульсная, Многоуровневая Квантизация), данные кодеки можно охарактеризовать, как комбинацию АЦП/ЦАП и вокодера. Применение вокодера позволяет снизить скорость передачи данных в канале. Основной принцип работы вокодера – синтез исходного речевого сигнала посредством адаптивной замены его гармонических составляющих соответствующим набором частотных фонем и согласованными шумовыми коэффициентами. Кодек G.723 осуществляет преобразование аналогового

сигнала в поток данных со скоростью 64 Кбит/с (ИКМ), а затем при помощи многополосного цифрового фильтра/вокодера выделяет частотные фонемы, анализирует их и передаёт по IP-каналу информацию только о текущем состоянии фонем в речевом сигнале. Данный алгоритм преобразования позволяет снизить скорость кодированной информации до 5,3 – 6,3 Кбит/с без видимого ухудшения качества речи. Кодек G.723.1 широко применяется в голосовых шлюзах и прочих устройствах IP-телефонии. Кодек уступает по качеству кодирования речи кодеку G.729 [8, 9], но менее требователен к ресурсам процессора и пропускной способности канала.

Заключение. Объективные критерии оценки качества воспроизведения речи используют формализованные параметры, позволяющие судить о степени различия переданного по каналу связи и восстановленного на приёмной стороне сигналов. Человек является основным потребителем услуг любой телекоммуникационной системы, поэтому качество сигнала может оцениваться по его субъективному восприятию речи. При выборе стандартных кодеков для передачи речи должно быть гарантировано именно качество воспроизведения речи.

Современные продукты для IP-телефонии применяют самые разные кодеки, стандартные и нестандартные. Конкурентами являются кодеки GSM (13,5 кбит/с) и кодеки МСЭ-Т серии G, использование которых предусматривается стандартом H.323 для связи по IP-сети. Единственным обязательным для применения кодеком в H.323-совместимых продуктах остается стандарт G.711.

В качестве дополнительных высокопроизводительных кодеков стандарт H.323 рекомендует G.723 и G.729 – последние способны сжимать оцифрованную 16-разрядную ИКМ-речь длительностью 10 мс всего в 10 байт. Стандарт G.729 уже получил широкое распространение в системах передачи голоса по каналам IP.

Список литературы: 1. Росляков А.В., Самсонов М.Ю., Шибяева И.В. IP-телефония. – М.: Эко-Трендз, 2003. – 252 с. 2. Жданов А.Г., Рассказв Д.А., Смирнов Д.А., Шитилов М.М., Передача речи по сетям с коммутацией пакетов (IP-телефония) / Под редакцией В.Ю. Бабкова и М.А. Вознюка. – С.-Пб.: СПбГУТ, 2001. – 148 с. 3. Антонян А.Б. Пакетная коммутация для передачи речи // Вестник связи. – 1999. – №5. – С. 68 – 71. 4. Ганьжа Д. Серия стандартов H.323 // LAN Magazine. – 1999, March. – С. 21 – 23. 5. Иавнова Т.И. Протокол H.323 в сетях IP-телефонии // Технологии и средства связи, 2000. – №2. – С. 94 – 99. 6. ITU-T: Recommendation G.711. Pulse code modulation (PCM) of voice frequencies. – Geneva, 1988. 7. ITU-T: Recommendation G.723.1. Dual rate speech coder for multimedia communication transmitting at 5.3 and 6.3 kbit/s. – Geneva, 1996. 8. ITU-T: Recommendation G.726. 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM). – Geneva, 1990. 9. ITU-T: Recommendation G.729. C source code and test vectors for implementation verification of the G.729 8 kbit/s CS-ACELP speech coder. – Geneva, 1996.

Поступила в редакцию 20.09.2004