**Как измерить качество речевой связи**

Питер Морриси

Передача речевой информации по IP-сетям, или IP-телефония (VoIP), предъявляет к сетевой инфраструктуре гораздо более жесткие требования, чем любое другое приложение. Степень доступности сети, потеря пакетов и временные задержки могут оказывать существенное влияние на качество речевой связи.

Чтобы исключить факторы негативного воздействия технических проблем в сети на качество связи, необходимо не только соответствующим образом спроектировать саму сеть, но и регулярно проводить мониторинг ее работы. Конечно, сама сеть передачи данных не является единственно возможным источником возникновения различных дефектов связи, проблемы могут также вызвать некачественные гарнитуры, непрофессионально сконфигурированные цифроаналоговые шлюзы и, наконец, тривиальное эхо из телефонной сети общего пользования (ТфОП). Поэтому вам придется выяснить, не оказывают ли эти компоненты отрицательное воздействие на работу речевых приложений.

Самым лучшим способом определения качества связи в сетях VoIP считается оценка его самими абонентами. Степень вредного воздействия шума, эха, задержек или помех от различных тоновых генераторов (warbling) на качество речевой связи определяется как интенсивностью этих факторов, так и тем, насколько они мешают разговаривать конкретным абонентам. Теоретически, для точной оценки степени этого воздействия следовало бы провести опрос достаточно большого числа абонентов, поскольку восприятие качества речевой связи всегда субъективно. Однако совершенно очевидно, что проведение таких регулярных опросов абонентов не представляется возможным.

Кроме того, нужно иметь в виду, что оценка степени удовлетворенности абонентов IP-телефонии качеством связи — дело достаточно сложное, в отличие, например, от подобной оценки качества услуги передачи данных, которое можно определить просто по скорости вывода на экран результатов тех или иных запросов (для этого достаточно лишь измерить время отклика системы). К счастью, появились новые средства контроля качества речевой связи, позволяющие делать это достаточно эффективно в автоматическом режиме.

**Старый, добрый MOS**

Официально рекомендуемым способом оценки качества IP- и других типов речевой связи является усредненная субъективная оценка MOS (Mean Opinion Score), алгоритм определения которой был разработана МСЭ-Т и последний раз корректировался в середине 90-х годов прошлого века. Алгоритм MOS, изложенный в спецификации МСЭ-Т Р.800, основан на пятибалльной шкале — от единицы (самое плохое качество связи) до пяти (самое хорошее качество). В соответствии с этим алгоритмом группа людей оценивает качество звучания тестовых речевых шаблонов, передаваемых через сеть. Эти оценки, конечно, субъективны. Как правило, самый большой балл, который можно получить по данной методике, равен 4,5. Рейтинг в 4,0 балла и выше рассматривается как очень высокий, соответствующий качеству ТфОП. Цифровые сеансы связи часто получают даже более высокие оценки.

Попытки автоматизировать процесс получения оценок MOS делались еще задолго до широкого распространения технологии VoIP. Один из таких алгоритмов, известный как PSQM (Perceptual Speech Quality Measure), был закреплен МСЭ-Т в спецификации Р.861. Однако PSQM, используемый в некоторых системах тестирования качества речевой связи, не слишком подходит для проверки VoIP-сетей, поскольку не учитывает, в частности, флуктуацию задержки (или джиттер), часто возникающую в IP-сетях. Между тем известно, что даже небольшая задержка между пакетами с речевой информацией вызывает изменения передаваемого речевого шаблона. В результате при сравнении отправленного и принятого шаблона может быть выставлена некорректная оценка качества речевой связи.

К счастью, современные программные средства, используемые для оценки качества речевой связи, выполняют свою работу максимально приближенно к тому, как это делает группа экспертов. Такие алгоритмы, как PSQM+, PESQ (Perceptual Evaluation of Speech Quality) и PAMS (Perceptual Analysis Measurement System), тоже предполагают передачу по VoIP-сети специальных речевых шаблонов и их последующий сравнительный анализ на приемной стороне сети. Алгоритм PESQ определен в стандарте МСЭ-Т Р.862.

**Разработчики программного обеспечения MOS**

Opticom (http://www.opticom.com) — разработчик алгоритмов контроля качества речи, а также портативных анализаторов Opera,

Psytechnics (httptf/www.psytechftics.com) — известна своими основанными на стандартах алгоритмами контроля качества речи и дополнительными продуктами типа Psyvoip.

Telchemy (http://wwwAekbemy.com) — разработчик программного обеспечения для контроля качества речи, используемого такими компаниями, как Acterna, Brix Networks, Finisar и др,

**Насколько точны алгоритмы оценки качества связи**

Если вы не уверены в том, что алгоритмы типа PESQ могут достаточно точно оценить качество речевой связи, попытайтесь оценить эффективность используемого вами инструментария. Если вы приобретаете программу, оценивающую значения MOS, обязательно попросите показать вам для сравнения результаты, полученные с помощью этой программы, и оценки MOS, выставленные реальными людьми (при использовании тех же самых шаблонов). Чем ближе MOS-оценки, полученные «ручным» и автоматизированным способами, тем точнее алгоритм.

«Тестируя алгоритм PESQ, мы постоянно сравнивали полученные с его помощью результаты с оценками экспертов», — рассказывает Майк Хол-лиер, директор по технологиям компании Psytechnics, которая принимала непосредственное участие в разработке PESQ. Этот алгоритм был предложен МСЭ-Т совместно компаниями Psytechnics, Opticom и SwissQual. Специалисты указанной международной организации сравнивали результаты измерения качества передачи речевой информации, полученные по алгоритму PESQ, с экспертными оценками, предоставленными компаниями AT&T и ВТ. Корреляция между двумя группами оценок, по словам Холлиера, составила 0,95 — очень хороший результат. И в декабре 2003 г. МСЭ-Т одобрил алгоритм PESQ в качестве стандарта. PESQ, который рассматривают в качестве преемника PSQM, более точно измеряет качество телефонной связи через IP-сети.

Вместе с тем недостатком PESQ и других новых алгоритмов является то, что они основываются на сравнении аудиопотоков на входе и выходе системы. Это так называемый мониторинг методом «активного тестирования». Но такой мониторинг, помимо того, что неизбежно добавляет трафик в сети, вызывает и ряд специфических проблем по организации тестирования. В частности, не всегда представляется возможность разместить измерительное оборудование на обоих концах канала связи. Предположим, вызов в вашу IP-сеть поступает из ТфОП, и вы не в состоянии проверить качество шлюза ТфОП, который может генерировать эхо и вызывать дополнительную задержку при перекодировании речевых сигналов.

Таким образом, алгоритмы «активного тестирования» позволяют эффективно проверить качество связи внутри вашей сети, но мониторинг любых вызовов, заканчивающихся вне ее, оказывается невозможным. Для решения этой проблемы была разработана спецификация Р.563, которая определяет алгоритм «пассивного мониторинга». Такой алгоритм тестирования не предполагает введение в сеть дополнительного трафика, и реализующее его программное обеспечение анализирует непосредственно передаваемые речевые потоки. При этом программа генерирует рейтинги MOS аналогично тому, как это делает группа экспертов.

Этот метод отличается от многих других тем, что для сравнения в нем не используются никакие специальные речевые шаблоны. Рейтинг качества, полученный с помощью Р.563, может быть с легкостью верифицирован группой экспертов, работающих параллельно. Корреляция между рейтингами Р.563 и рейтингами, выставляемыми реальными пользователями, обычно варьируется от 0,85 до 0,9. Конечно, алгоритм Р.563 не так точен, как PESQ, но при этом он достаточно эффективен.

**ДРУГИЕ АЛГОРИТМЫ ОЦЕНКИ КАЧЕСТВА**

Компания Psytechnics предложила еще один алгоритм оценки качества — Psyvoip. Он менее точен, чем Р.563, зато способен анализировать речевой поток в реальном времени в течение всего сеанса связи. Это очень интересный вариант, особенно учитывая то, что для анализа только одного звонка по методике MOS требуется достаточно большая процессорная мощность. Алгоритм Psyvoip для оценки качества связи использует статистические данные протокола RTP (Real Time Protocol). Он также обнаруживает проблемные IP-адреса, чтобы потом с помощью более сложных алгоритмов, типа PESQ, проводить дополнительное исследование.

Еще один способ анализа качества речевой связи основан на модели E-model (стандарт G.107), изначально разработанной с целью облегчить планирование инсталляций VoIP-систем. Компания Qovia и ряд других разработчиков усовершенствовали этот алгоритм и адаптировали его для пассивного мониторинга. По точности оценки этот метод уступает Р.563, зато его реализации не так дороги. Системы, основанные на E-model, целесообразно использовать для локализации проблем и последующего их исследования с помощью более дорогих и ресурсоемких алгоритмов.

**Глоссарий**

Active monitoring — метод оценки качества речевой связи, предполагающий посылку по сети специального контрольного речевого шаблона с последующим сравнением его на входе и на выходе системы и оценкой степени воздействия сети на этот шаблон. Иногда, этот метод называют мониторингом с вмешательством в работу сети (intrusive monitoring).

E-Model (МСЭ-Т G.107) — алгоритм, первоначально предназначавшийся для использования при планировке сети, был доработан некоторыми компаниями с целью применения в качестве недорогого инструментария пассивного мониторинга. Оценивает качество речевой связи с помощью R-фактора (по шкале от 0 до 100).

MOS (Mean Opinion Score) — алгоритм, закрепленный в спецификации МСЭ-Т Р.800, предполагает систематическую экспертную оценку качества голосовой связи группой пользователей по пятибалльной шкале. Оценки 4,0 и выше рассматриваются как соответствующие качеству связи в ТфОП.

PAMS (Perceptual Analysis Measurement System) — нестандар-тизованный алгоритм оценки качества голосовой связи, разработанный компанией British Telecom в 1998 г. Учитывает флуктуации задержек в cети VoIP.

Passive monitoring — предполагает контроль качества речевой связи путем прослушивания разговора на линии и последующей оценки качества сигнала без сравнения с каким-либо контрольным шаблоном.

PESQ (Perceptive Evaluation of Speech Quality) — алгоритм оценки качества речевой связи, закрепленный МСЭ-Т в спецификации Р.862. Считается самым современным и эффективным алгоритмом оценки качества голосовой связи в сетях VoIP. Использует активные способы тестирования.

Р.563 — спецификация МСЭ-Т, утвержденная в мае 2004 г. Определяет алгоритм пассивного мониторинга для оценки качества речевой связи путем прослушивания «живых» сеансов связи.

PSQM (Perceptual Speech Quality Measure) — алгоритм МСЭ-Т Р.861.

Используется для тестирования кодеков. Неприменим для тестирования VoIP-систем, поскольку не учитывает возникающие задержки при передаче голоса и их флуктуации.

PSQM+ — улучшенная версия алгоритма PSQM, разработанная компанией Qpticom специально для оценки систем VoIP. Пока не стандартизована.

RTCP (Real Time Control Protocol) — протокол, предполагающий, в частности, посылку сообщений со статистикой о качестве связи от одной конечной точки VoIP до другой. Не все производители реализовали его в своих оконечных устройствах.

RTCP-XR (RFC3611) — улучшенная версия протокола RTCP, предполагает обмен дополнительной информацией о качестве связи — например, о резком увеличении числа потерянных пакетов. Этот протокол является опциональным, и его поддерживают не все производители.

RTP (Real Time Protocol) — протокол передачи аудиофайлов по сетям передачи данных, обычно по протоколу UDP.

Так какой же алгоритм является оптимальным для большинства компаний? Первое место несомненно за PESQ, самым предпочтительным алгоритмом активного мониторинга. Вслед за ним идет Р.563, единственный стандартизованный алгоритм пассивного мониторинга. Некоторые компании считают, что в ряде случаев могут быть достаточно эффективны недорогие расширения технологии G.107. Убедитесь, что вы хорошо понимаете ограничения, присущие каждому из предлагаемых вам алгоритмов. Кроме того, еще до покупки тщательно изучите данные о корреляции результатов автоматизированных и «ручных» тестов.

**Как найти оптимальный алгоритм**

Приобрести продукты Psytechnics и Opticom обычным способом невозможно, потому что их используют п основном компании, производящие системы сетевого управления и контрольно-измерительную аппаратуру, такие, как Agilent Technologies, Brix Networks, Empirix, Radcom и Qovia. Они, в свою очередь, предлагают их уже в комплекте со своими продуктами. В результате вы получаете возможность пользоваться алгоритмами MOS, а также систематически собирать раз личные данные для генерации отчетов и аварийных сообщений. В некоторых случаях оказывается возможным даже коррелировать данные анализа качества речевой связи с другими событиями в сети, используя сетевую статистику, что существенно облегчает диагностику возникающих- проблем. Некоторые продукты перечисленных выше компаний генерируют отчеты для проверки соглашений об уровне обслуживания SLA (Service Level Agreements).

При более детальном ознакомлении с программными средствами генерации оценок MOS выясните, какие проблемы они позволяют локализовать, в каком формате выдают информацию и как ее следует интерпретировать. Например, несмотря на то, что эти средства позволяют получить довольно точную оценку качества связи, они не всегда помогут вам выявить источник возникшей проблемы с качеством связи. А не зная причины, вы не сможете устранить возникшую проблему. Также нужно отдавать себе отчет, что выдаваемый некоторыми продуктами MOS-рейтинг представляет собой оценку качества некоторого среднего вызова, а качество конкретных сеансов связи может существенно отличаться от рейтинга MOS.

Помимо тестирования своей собственной сети, вы можете использовать эти продукты и для оценки уровня сервиса своего поставщика услуг VoIP. Любому пользователю гораздо удобнее ориентироваться на конкретные показатели качества, оговоренные в SLA, чем на параметры типа число потерянных пакетов, уровень джиттера или значение задержки. Другими словами, дело сервис-провайдера — обеспечивать нормальную работу сети, а в ваши функции должен входить лишь регулярный ее мониторинг, позволяющий удостовериться в том, что вам обеспечен именно тот уровень обслуживания, за который вы платите деньги. (Не лишним будет поинтересоваться у вашего потенциального или существующего сервис-провайдера, какими средствами он пользуется для проак-тивного поиска возможных проблем).

Кроме мониторинга качества ТР-телефонной связи, вам, возможно, захочется отслеживать и другие параметры — например, время установки VoIP-соединения. Если от момента набора номера до момента установления связи проходит более 10 с, скорее всего, ваши пользователи будут не довольны качеством связи. Это становится большой проблемой в работе call-центров, где время установки соединения также важно, как и качество связи. Одно программное обеспечение MOS не в состоянии решить описанную задачу, но если его использовать в комплекте с другими продуктами, например с системой мониторинга компании Qovia, которая отслеживает не только качество связи, но и любые изменения б работе IP-УАТС, то решение этой задачи вполне реально.

**Поставщики систем управления и контроля, использующих алгоритм MOS**

Brix Networks (http://www.brixnetworks.c0m) - разрабатывает средства проверки SLA и управления характеристиками сети для крупных организаций и операторов связи»

Empirix {http://www.empirix.com) — выпускает продукт» Hammer дня тестирования и мониторинга качества систем VoIP-связи» а также качества работы call-центров.

Qovia {http://www.qovia.com) — предлагает системы сетевого управления для мониторинга качества IP-телефонной связи, а также других связанных с качеством параметров в сетях передачи данных. Radcom (http://www.radcom.com) — разрабатывает системы сетевого управления и средства тестирования для VoIP- и сотовых сетей связи. Spirent Communications (http://www.spirentcom) — известна своими продуктами для тестирования VoIP-сетей, такими, например, как Abacus 6000.

**Список литературы**

«Сети и системы связи» № 08 2005 г.