

## Оптимизация мобильного трафика в сетях 3G и LTE

При всей своей распространенности и широте использования приложениями, работающими на мобильных абонентских устройствах, протокол TCP был и остается оптимизированным только для работы в проводных сетях. Полагаясь на надежность нижележащих уровней, присущую проводным сетям, TCP предполагает, что любая потеря пакетов данных вызвана сетевыми перегрузками, и начинает предпринимать меры по их предотвращению.

**А.Фелижанко,**  
компания Cisco systems

Базовый механизм борьбы с перегрузками в TCP основан на управлении размером так называемого окна перегрузки, то есть тем количеством данных, которое отправитель передает в сеть за один раз. На начальной фазе TCP "прощупывает" полосу пропускания, доступную для соединения. На этой фазе работает алгоритм медленного старта, при котором размер окна перегрузки от начального значения, равного размеру одного максимального сегмента, до установленного порогового значения растет экспоненциально, а затем, на фазе предотвращения перегрузок, линейно. При наступлении таймаута, когда подтверждение приема пакета данных получателем не приходит в отведенный срок, новое пороговое значение устанавливается на половину текущего окна перегрузки, само окно перегрузки схлопывается до начального значения, и работа алгоритма начинается сначала. Если таймауты больше не наблюдаются, размер окна перегрузки вырастает до размера окна, объявленного получателем, и отправитель начинает передавать в сеть количество данных, соответствующее размеру окна получателя. И так до очередного таймаута. Очевидно, что чем чаще срабатывает базовый механизм, тем больше провалов в скорости передачи и тем менее эффективно используется доступная для соединения полоса пропускания.

В беспроводных каналах связи кратковременные периоды потерь пакетов или резкого роста задержки их доставки, приводящие к наступлению таймаутов TCP, могут быть вызваны причинами, не связанными с реальными перегрузками. Среди этих причин: затухание радиоканала, помехи и другие преходящие условия, характерные для

беспроводной среды; мобильность абонентов и связанные с ней временные потери радиоканала, например, при проезде через туннель или пользовании лифтом; перемещение абонентов между соседними базовыми станциями; повторные передачи пакетов протоколами канального уровня; настройки промежуточных узлов, например, eNodeB, на приоритетное обслуживание определенных услуг и предоставление им лучшего качества обслуживания, в результате чего остальные услуги могут испытывать вариацию задержки. Эти спорадические события вызывают включение достаточно консервативного базового механизма управления перегрузками TCP, что приводит к заметному снижению эффективности его работы и негативно сказывается на фактической пропускной способности: растет время открытия веб-страниц, скачивания файлов и т.д., а полоса пропускания радиосети используется не полностью.

Одним из выходов из этой ситуации может быть установка некоего посредника, или прокси, который разбивает соединение TCP между клиентом на абонентском устройстве и сервером TCP в Интернете на две независимые части, условно говоря, "беспроводную" и "проводную", и более интеллектуально управляет перегрузками на беспроводной части соединения. Поскольку основной объем трафика в Интернете передается в направлении от сервера к клиенту, на беспроводной части соединения в сторону абонента отправитель TCP такого посредника волен применять любые алгоритмы управления размером своего окна перегрузки. Получателю TCP на абонентском устройстве до этого нет никакого дела, он ведет себя обычным образом, и каких-либо модификаций стандартного стека TCP на нем не требуется.

Именно на этих принципах основана реализация функциональности Traffic

Performance Optimization (TPO) на многофункциональной платформе Cisco ASR 5000, когда она выступает в качестве шлюзового узла GGSN сети 3G или в качестве шлюза PDN Gateway (PGW) пакетного ядра сети LTE. На проводной части соединения между TCP-прокси на GGSN/PGW и сервером TCP в Интернете работает обычный протокол TCP без каких-либо изменений. А на беспроводной между TCP-прокси и клиентом TCP на абонентском устройстве - оптимизированный TCP. Оптимизация TCP затрагивает различные аспекты протокола, начиная от процедур установления соединения и медленного старта до управления перегрузками в различных ситуациях. Так, для управления размером окна перегрузки и выбора порогового значения для медленного старта после наступления таймаута используются методы TCP Reno, TCP Vegas или TCP Westwood+. Эти методы принимают в расчет динамику изменения времени прохождения пакетов в прямом и обратном направлениях (RTT), а также оценку доступной для соединения полосы пропускания (как результат мониторинга интенсивности получения подтверждений на отправку данных), по различным признакам пытаются предугадать момент наступления перегрузки, чтобы предпринять превентивные меры, а также понять причину потери пакета - действительно ли это результат перегрузки или следствие аномалий в беспроводном канале. В дополнение, для более тонкого управления перегрузками используются эвристические алгоритмы, учитывающие параметры качества обслуживания абонента, в частности, величину гарантированного битрейта (GBR) на соединении. Кроме того, работа алгоритмов корректируется при получении шлюзом PGW извещений о хэндовевере абонента между базовыми станциями с одной и той же технологией, например, из LTE в LTE, и с различными технологиями, например, из

LTE в 3G. Лабораторные исследования показывают, что на беспроводном соединении с задержкой 100 мс и полосой пропускания 1 Мбит/с при доле потерь пакетов 0,5% фактическая пропускная способность канала при работе ТРО по сравнению с базовой увеличивается более чем на 30% и приближается к теоретической. При доле потерь пакетов 1% она увеличивается почти на 60%.

Другой аспект работы ТРО на Cisco ASR 5000 связан с оптимизацией трафика HTTP. Веб-страницы зачастую не адаптированы к условиям беспроводных сетей, приложения на абонентских устройствах делают множество служебных запросов к серверам DNS в Интернете для разрешения доменных имен в адреса IP, серверы контента не всегда понимают, что запросы идут от мобильных устройств, и не подстраивают контент соответствующим образом. Логичным было бы оптимизировать трафик HTTP на GGSN или на PGW. Именно это и реализовано на Cisco ASR 5000. Если контент на веб-сервере текстовый, а браузер на абонентском устройстве способен принимать компрессированные данные, ТРО будет сжимать содержимое страниц HTML и передавать их в сторону абонента в формате gzip. Как показывает практика, экономия полосы пропускания в радиоканале от такой компрессии составляет в среднем 30-40%. Это с учетом того, что формат JPEG уже сжат: текстовые страницы без JPEG сжимаются до 70%. Если страница HTML в ответе веб-сервера содержит ссылки на вспомогательные ресурсы, например, на изображения JPEG, ТРО будет сам разрешать доменные имена в ссылках URL и переписывать их на адреса IP с тем, чтобы минимизировать количество обращений браузера на абонентском устройстве к серверу DNS. Если же в страницу встроена реклама в виде изображения или Flash-объекта, ТРО будет ее блокировать с возможностью последующей загрузки по запросу (on-click). Эффективность этих методов во многом зависит от сайта, к которому обра-

щается абонент. При включенной оптимизации HTTP время открытия страниц, например, CNN.com сокращается почти в полтора раза, количество запросов DNS со стороны абонента — на треть, а эффект от компрессии страниц достигает 35%.

Еще один аспект оптимизации использования радиоресурсов связан с мобильным видео, доля которого в общем объеме мобильного трафика к 2015 г. будет составлять две трети. Это данные исследования компании Cisco "Прогноз развития мирового мобильного трафика на 2010-2015 гг.". Для распознавания видеозапросов со стороны абонентов шлюз мобильного видео на Cisco ASR 5000 использует глубокую инспекцию пакетов. Эти запросы могут быть переадресованы на вспомогательный сервер адаптации видеоконтента, причем шлюз может выполнять балансировку видеозапросов HTTP между серверами адаптации, входящими в кластер. Сервер адаптации контента передает в сторону абонента уже оптимизированное, например, транскодированное видео. На шлюзе также отработываются правила обработки видеотрафика, в которых может быть указано, какой трафик и для каких абонентов подлежит оптимизации средствами самого шлюза, в частности, динамическому трансрейтингу и пэйсингу, а какие видеоклипы следует пропускать в исходном виде. Логично, что функциональность шлюза мобильного видео в сети 3G включается на GGSN, а в сети LTE — на PGW. Она дополняет ТРО и автоматически активирует функцию проксирования и оптимизации TCP, в том числе для оценки доступной полосы пропускания на беспроводной части соединения.

Трансреитинг в реальном масштабе времени дает возможность снижать видео-битрейт прогрессивной загрузки видео до величины, поддерживаемой конкретной беспроводной сетью. Например, с 25 и 30 фреймов в секунду (соответственно, для PAL и NTSC) до 8, 12 или 15, достаточных для воспроизведения видео на мобильном устрой-

стве с относительно небольшим экраном без сколько-нибудь заметного ухудшения качества, но со значительной экономией полосы пропускания в радиосети. С другой стороны, видеоэйсинг позволяет привести в соответствие скорость воспроизведения видео на мобильном устройстве со скоростью его загрузки. Речь идет о просмотре видеоклипов, прогрессивно загружаемых и просматриваемых через HTTP, например, с YouTube. Как правило, видео сначала загружается в буфер устройства, а затем при достаточном его заполнении начинает проигрываться. Если скорость загрузки превышает скорость воспроизведения, буфер в процессе просмотра заполняется полностью. Однако по статистике просмотров короткого (порядка 4 мин) видео, характерного для YouTube, 20% зрителей прекращает просмотр видеоклипов через 10 или менее секунд, треть через 30 секунд, 44% через 1 минуту, а уже 60% всего через 2 минуты просмотра. Как следствие, полоса пропускания в радиосети расходуется впустую. Видеоэйсинг же позволяет значительно сократить объем видеотрафика. Опытная эксплуатация шлюза мобильного видео на Cisco ASR 5000 в сети одного из ведущих операторов мобильной связи США подтвердила, что только за счет видеоэйсинга объем видеотрафика в радиосети сократился на 26%, а общий объем мобильного трафика, в состав которого входил и видеотрафик, — на 6%. И все это без каких-либо нареканий на качество видео со стороны абонентов.

В заключение хотелось бы отметить, что в результате работы описанных методов оптимизации мобильного трафика у абонентов создается лучшее впечатление о качестве предоставляемых услуг, отодвигаются сроки или вообще ставится под сомнение целесообразность дополнительных инвестиций в немедленное увеличение мощности радиосети, что, как правило, является первой реакцией оператора на появление пеленгровок и жалоб со стороны абонентов.

