

*ДЕМЧИНСКИЙ В.В.,
ДОРОГОЙ Я.Ю.,
ДОРОШЕНКО К.С.*

АНАЛИЗ И ПРАКТИКА ВНЕДРЕНИЯ МЕХАНИЗМОВ КАЧЕСТВА ОБСЛУЖИВАНИЯ

В статье рассмотрены вопросы, связанные с анализом и практикой внедрения механизмов качества обслуживания.

The questions related to the analysis and the introduction of quality of service was described in this article.

Введение в проблему

Важность качества предоставляемых услуг для успешного ведения бизнеса очевидна. Рассмотрим одну из сторон этого вопроса в сфере телекоммуникационных услуг - обеспечение требуемых характеристик сервиса для определенных классов трафика. Будем называть трафиками потоки данных по тем или иным признакам выделяемые из совокупного потока данных. Особенность современной сетевой инфраструктуры состоит в совместной передаче трафиков различной природы (видео, голос, данные) по единой сети (конвергенция, TriplePlay). Для каждого типа трафика можно обозначить типичные характеристики и требования, обуславливаемые особенностями приложений, генерирующих данные трафики. Определим понятие качество обслуживания (QoS, Quality-of-Service) как способность сетевых средств обеспечивать требуемые характеристики сервиса для определенных классов трафика. Интересующие характеристики передаваемого трафика - задержка, вариация задержки и потери пакетов.

Основное назначение технологий качества обслуживания состоит в обеспечении приоритетов, необходимой полосы пропускания, контролируемой задержки и вариации задержки, уменьшении процента потерь пакетов при передаче, а также в обеспечении приоритетности обслуживания некоторых потоков с возможностью одновременной передачи других потоков. Зачастую (из-за мультиплексирования трафика или из-за превышения ПС (пропускной способности) входного канала над ПС выходного) сетевые устройства не успевают передавать весь поступающий трафик, что вызывает накопление пакетов в очередях и сброс пакетов при исчерпании доступной памяти.

Под перегрузкой будем понимать такое состояние сети, при котором основные показате-

тели качества обслуживания заметно ухудшаются. В технологиях QoS применяются 2 стратегии: управление перегрузками и предотвращение перегрузок. Управление перегрузками состоит в управлении очередями, а предотвращение перегрузки осуществляется механизмами контроля за установлением соединений, формирования трафика и управления скоростью передачи.

Цель работы

Целью работы является исследование параметров качества предоставления услуг и их зависимости от применения того или иного типа очередей.

Архитектура QoS

Архитектура QoS [1] включает механизмы классификации, маркировки и формирования (профилирования) трафика, управления очередями и сигнализации QoS (метки DSCP, распространение правил QoS по протоколу BGP, резервирование ресурсов RSVP). В определенном смысле, технология качества обслуживания №1 - пропускная способность, т.е. увеличение ПС безусловно улучшает характеристики QoS. Однако запросы сетевых приложений эволюционно растут и практически при любой доступной ПС несложно будет ее утилизировать. Т.е. само по себе наращивание ПС (без внедрения других механизмов QoS) будет только временным решением и способно даже усугубить ситуацию. "Возможно купить пропускную способность, но нельзя купить задержку".

В зависимости от требований приложений и условий соглашения трафик подвергается дифференциации по нескольким уровням приоритета или классам обслуживания. Как правило, механизмы классификации, маркировки и формирования трафика применяются

как можно ближе к границе сети. Классификация трафика на границе сети позволяет освободить ядро (магистраль) сети от этой задачи. Пакеты могут быть классифицированы либо устройством на входе в сеть, либо же породившим трафик приложением. Способы классификации данных можно разделить на явные (метки, заданные в пакетах отправителем) и неявные (выбираемые сетевыми устройствами исходя из заданной политики). Пакеты могут быть классифицированы по любым параметрам заголовков (в общем случае - по спискам контроля доступа) или используя механизм распознавания сетевых приложений.

Как правило, качество обслуживания необходимо обеспечивать как на границе кампусной сети, так и на магистральных соединениях [2]. Все механизмы качества обслуживания должны работать согласованно, поскольку несогласованная работа одного единственного узла сети может нивелировать весь результат. В свете этого полезно исследование поведения трафика при совместной работе маршрутизаторов с поддержкой QoS и без таковой (единственная очередь FIFO).

Заявляемые характеристики обслуживания называют Соглашением об уровне сервиса (Service Level Agreement, SLA) Обычно соглашение об уровне сервиса включает параметры пропускной способности, задержки, потерь пакетов, время реакции/восстановления и условия обслуживания. В целом, требования необходимого уровня QoS определяются решаемыми задачами и используемыми приложениями. При планировании внедрения QoS [3] следует учитывать особенности каждого типа трафика, в первую очередь, трафика реального времени. Интерактивный трафик наиболее чувствителен к параметрам задержки и вариации задержки. Обычно для комфортной работы интерактивных приложений передачи голоса или видео полная задержка не должна превышать 150 мс, а вариация задержки - до 10..30 мс [4]. Для потоковой передачи в реальном времени требования менее жесткие, соответственно 1 с и 100 мс. Суммарная (коммулятивная) задержка передачи состоит из задержек сериализации, прохождения сигнала по среде передачи и задержки ожидания в очередях. Вариация задержки определяется размерами пакетов других приложений. Рассмотрим характеристики основных механизмов управления очередью [1,2]. Аппаратно практически все механизмы управления очередями работают между

буфером интерфейса и очередью на отправку (передающим кольцом интерфейса).

Общая очередь FIFO. Пакеты всех потоков рассматриваются как равноправные и обслуживаются по мере их поступления. По сути, механизм FIFO не предполагает никакого активного управления трафиком. Быстрота обработки пакетов в очередях зависит только от очередности их поступления и скорости обработки.

Взвешенная справедливая очередь (Weighted Fair Queue, WFQ). Весь трафик одного уровня приоритета попадает в классовую очередь, в которой все потоки обрабатываются равномерно и получают приблизительно равную задержку. Предполагается, что потоки с меньшим объемом данных будут иметь приоритет над потоками с большим объемом данных. При увеличении количества потоков доля полосы каждого потока будет уменьшаться. Считается, что планировщик WFQ улучшает работу протокола TCP (медленный старт и контроль за перегрузкой). Это приводит к более равномерной загрузке каналов и стабильному времени ответа TCP приложений. Существуют механизмы взвешенной справедливой очереди, основанные на потоках или на классах. В первом случае сетевые устройства автоматически распределяют трафик на потоки и обеспечивают равномерное распределение ПС между ними. Во втором случае классификация трафика является настраиваемой. При большом количестве обрабатываемых потоков механизм WFQ увеличивает нагрузку на процессор сетевого устройства, поэтому его применение на высокоскоростных интерфейсах ограничено. Взвешенная справедливая очередь позволяет динамически разделить полосу пропускания между потоками согласно весу каждого потока, определяющего порядок обработки пакетов в очереди, и дает возможность снижать задержки потоков трафика.

Взвешенная справедливая очередь на базе классов (Class-Based Weighted Fair Queue, CBWFQ) позволяет обеспечивать минимальную гарантированную полосу пропускания и предотвращает захват полосы пропускания низкоприоритетными потоками при передаче высокоприоритетных данных. Если один из классов не полностью использует заданную долю ПС, то она равномерно распределяется между другими потоками. Аппаратно взвешенная справедливая очередь реализуется механизмом очередности с весами и с учетом

длины пакетов. По сути, для всех технологий QoS должен выполняться закон сохранения задежки.

Приоритетная очередь (Priority Queue, PQ). Механизм PQ обеспечивает безусловный приоритет трафика. Реализуется 4 очереди со строгим приоритетом. Приоритетные очереди целесообразно применять лишь в тех случаях, когда трафик имеет иерархическую структуру и трафики с более высоким приоритетом не должны задерживаться нижними трафиками. Недостаток приоритетной очереди – пагубное воздействие на неприоритетные трафики и непредсказуемость их времени ожидания. Потoki данных, требующие большой ПС, не должны попадать в высокоприоритетную очередь, а непродуманная настройка приоритетов может привести к блокированию низкоприоритетных трафиков.

Настраиваемая очередь (Custom Queue, CQ). Планировщик CQ делит ПС между потоками пропорционально их весу. Очереди обслуживаются циклически, из каждой очереди берется число байт пропорционально весу очереди. ПС канала в результате динамически распределяется между очередями. Выделяемая очереди полоса пропускания задается неявно посредством счетчика байт и длины очереди. Настраиваемые очереди используются в тех случаях, когда необходимо гарантировать определенные минимальные характеристики для каждого класса трафика (неявно резервировать полосу пропускания) и не дают возникнуть ситуации, когда пакет будет находиться в очереди в течение неопределенного времени. Настраиваемые очереди позволяют более равномерно обслуживать различные потоки и, по сравнению с приоритетными очередями, не блокируют низкоприоритетный трафик. Аппаратно настраиваемые очереди реализуются с помощью механизма весов с заимствованием. Обычно параметры настраиваемых очередей задаются статически и не могут автоматически адаптироваться к изменению ситуации в сети.

Очередь с малой задержкой (Low Latency Queue, LLQ) - комбинация приоритетной очереди и классовой взвешенной справедливой очереди. В очереди с малой задержкой для определенного класса резервируется минимальная полоса пропускания, а в случае доступности дополнительных ресурсов класс способен их утилизировать. Если класс не использует гарантируемую полосу - ее могут использовать другие потоки.

Формирование трафика - процедура манипулирования данными для повышения качества передачи. Средства выполнения правил (policing) сбрасывают пакеты, выходящие за пределы оговоренной скорости, а средства профилирования пытаются буферизировать и сгладить выбросы трафика, и таким образом создать равномерный поток. Механизм профилирования трафиков требует больше ресурсов и применяется, как правило, на пограничных клиентских устройствах, в то время как механизм выполнения правил – между сетями.

Эксперименты и результаты

Рассмотрим примеры поведения трафика при использовании различных механизмов QoS. Смоделируем следующую топологию: каналы Ethernet с ПС 10 Mbit/s, последовательные каналы - 800 Kbit/s. Для наглядности сравнения относительных характеристик трафика возьмем 4 потока с равной интенсивностью. Средняя скорость каждого потока 44 пакета/секунду, длина пакетов равномерно распределена от 200 до 1000 байт. Интенсивность трафика подобрана таким образом, чтобы наиболее узкий канал в топологии (800 Kbit/s) постоянно находился в состоянии насыщения (минимально возможные потери). При моделировании использовался генератор трафика TGN [5]. Заданный поток трафика передается от R4 через R1, R2, R3, R5 и поступает на другой порт R4, где и происходит сбор его статистик (рис. 1).

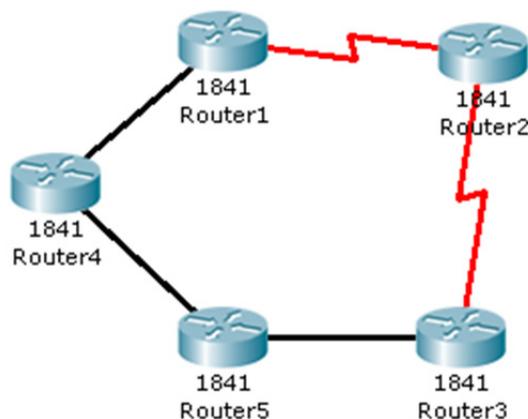


Рис.1. Топология сети

Будем изменять механизм управления очередью на выходном интерфейсе маршрутизатора R1. На остальных маршрутизаторах - очереди FIFO. Эксперименты показали, что повторное применение механизмов QoS к сформированному трафику с заданными характеристиками не улучшает его характеристики. Понаблюдаем за метриками трафика в

зависимости от применяемых на маршрутизаторе R1 механизмах QoS. Интерес представляет относительное изменение метрик при разных сценариях моделирования. Базовый сценарий (сценарий 1) - отсутствие управления трафиком на R1 (общая очередь FIFO). Сценарий 2 - использование WFQ очереди. Сценарий 3 – использование CBWFQ очереди, веса всех 4-х классов равны. Сценарии 4 и 5 – CBWFQ очередь, в которой веса 1-го класса увеличены, соответственно, до 0.30 и 0.35 (за счет уменьшения веса 4-го класса трафика). Сценарий 6 - очередь LLQ, веса всех классов равны, 1-й трафик приоритетный.

Результаты моделирования (табл. 1) свиде-

детельствуют, что за счет организации очередей в планировщике WFQ (сценарий 2) немного увеличилась задержка, но уменьшился разброс (вариация) задержки - трафик стал более предсказуемым.

Сценарии 3, 4 и 5 показывают - увеличение доли потока можно рассматривать как увеличение относительного приоритета, что влечет уменьшение задержки данного потока за счет увеличения задержки "класса по-умолчанию". Сценарий 6 - использование приоритетной очереди для 1-го класса трафика значительно уменьшает его задержку (практически до минимально возможного уровня - постоянной составляющей задержки) и перераспределяет

стью подавить остальные трафики.

Предпочтительнее использование LLQ очереди, где для приоритетного трафика ограничена максимальная доля используемой ширины канала. В этом случае также следует оценивать максимально возможную интенсивность приоритетного трафика, чтобы обеспечить ему достаточную полосу пропускания. Следовательно, для максимального использования технологий QoS целесообразно формирование характеристик трафика на входе в сеть и применение скоростных механизмов управления очередями на магистральных соединениях.

Выводы

Таким образом, существующие механизмы управления очередями способны обеспечивать равномерное обслуживание различных потоков трафика и приоритет их обслуживания. Относительный приоритет трафика может быть задан как вес потока или посредством реализации отдельной очереди потока.

При абсолютной приоритетности трафика обязательно следует учитывать его максимально возможную интенсивность т.к. трафик с абсолютным приоритетом способен полно-

Табл. 1. Результаты моделирования

Сценарий	1		2		3		4		5		6	
	avg-delay	stdev-delay										
1	44,58	8,70	39,64	19,67	54,53	11,54	40,49	10,05	14,28	6,93	2,13	0,86
2	47,01	8,81	40,03	19,55	54,15	10,73	54,12	11,09	53,75	10,84	63,71	14,35
3	45,74	8,86	39,29	19,69	55,01	11,46	54,17	10,11	53,40	11,00	64,97	15,93
4	46,31	7,38	39,68	19,69	54,64	10,26	69,48	12,94	90,37	17,84	63,45	15,43

Список литературы

1. Кучерявый Е.А. Управление трафиком и качество обслуживания в сети Интернет. - СПб.: Наука и Техника, 2004. - 336 с.: ил.
2. Вегешна Шринивас. Качество обслуживания в сетях IP. :Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 368 с.
3. Степанов С.Н. Основы телетрафика мультисервисных сетей. - М.: Эко-Трендз, 2010. - 392 с.: ил.
4. Девидсон Джонатан и др. Основы передачи голосовых данных по сетям IP. 2-е издание. :Пер. с англ. – М.: Издательский дом «Вильямс», 2007. – 397 с.
5. TGN Traffic Generation Using Cisco Pagent IOS. – [Электронный ресурс]. – URL: <http://www.ciscoconsole.com/nms/traffic-generation-using-cisco-pagent-ios.html/>