

УДК 004.724.4

А.С. Крупский, И.В. Бойченко

Алгоритм планирования QoS с учетом фрагментации пакетов в сетях WiMax

Рассматриваются известные алгоритмы планирования EDF и WFQ в условиях возникновения фрагментации пакетов в сетях WiMax. Предложены модификации этих алгоритмов, позволяющие компенсировать потери полосы пропускания. Дано сравнение реализаций модифицированных алгоритмов с их эталонным вариантом по критерию соблюдения качества обслуживания (QoS). Приведена методика оценки потерь полосы за счёт наличия фрагментации пакетов.

Ключевые слова: качество обслуживания, планирование ресурса, фрагментация, накладные расходы.

Задача обеспечения качества сервиса в сетях WiMax

Большинство существующих компьютерных сетей используют принцип Best Effort для распределения полосы пропускания среди абонентов. При этом качество предоставляемых услуг передачи не гарантировано, и трафик занимает всю свободную полосу. Тем не менее существуют такие сервисы (трансляция видеопотока, осуществление управления машинами и агрегатами, производственными системами в реальном времени с заданным временем отклика), которые чувствительны к изменениям характеристик каналов. Для обслуживания таких сервисов были созданы протоколы RTP [1] и RTCP [2]. Первый позволяет компенсировать дрожание полосы (англ. jitter) и обнаруживать нарушение последовательности пакетов. Второй служит для определения QoS (англ. Quality of Service – качество обслуживания), обеспечения обратной связи и синхронизации потоков получателя и отправителя. То есть осуществляется попытка компенсировать ухудшение качества при передаче через протокол транспортного уровня (на деле, чаще RTP и RTCP реализуются поверх UDP, так как гарантия доставки в TCP вызывает дополнительные задержки). Для обеспечения гарантии QoS на основе RTP, RTCP и RSVP [3] были введены модели Integrated Service [4] и Differentiated Service [5] для работы на сетевом уровне. Далее будет рассмотрена задача соблюдения QoS одним отдельно взятым узлом на канальном уровне, на основе спецификации беспроводной сети WiMax [6].

Постановка задачи

Задача обеспечения QoS на узле может быть представлена как задача оптимизации: максимизировать количество передаваемых данных при заданных ограничениях QoS для сервисных потоков. Ограничения должны учитываться только в том случае, если у абонента имеются данные для отправки. Стандарт WiMax выделяет следующие типы потоков и списки параметров QoS (табл. 1):

Таблица 1

Типы потоков и их параметры QoS

Приоритет. Тип	Список параметров QoS	Предполагаемые типы трафика
UGS(Unsolicited Grant Service)	Maximum Sustained Traffic Rate (P_{max_i}) Maximum Lattency (L_{max_i}) Tolerated Jitter (J_{max_i})	Применяется для сервисов, требующих фиксированной во времени полосы (T1/E1 and Voice over IP)
ERT(Extended Real Time)	Maximum Sustained Traffic Rate(P_{max_i}) Minimum Reserved Traffic Rate(P_{min_i}) Maximum Latency(L_{max_i})	Сочетает в себе UGS и RT – фиксированную полосу в задаваемые клиентом промежутки времени
RT(Real Time)	Maximum Sustained Traffic Rate (P_{max_i}) Minimum Reserved Traffic Rate (P_{min_i}) Maximum Latency(L_{max_i})	Предназначен для сервисов, требующих выделения изменяющейся во времени полосы на регулярной основе (MPEG)
NRT(Not Real Time)	Maximum Sustained Traffic Rate(P_{max_i}) Minimum Reserved Traffic Rate (P_{min_i})	Предназначен для сервисов, требующих постоянной полосы в среднем (FTP)
BE(Best Effort)	Maximum Sustained Traffic Rate (P_{max_i})	Сервисы, обслуживающиеся по возможности, стандарт не предоставляет гарантий

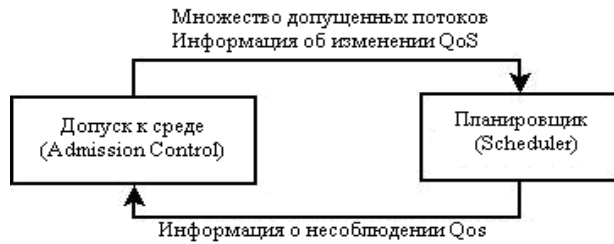


Рис. 1. Архитектура системы планирования

Архитектура системы планирования

Для решения задачи планирования на станциях WiMax стандарт [6] предлагает архитектуру, состоящую из двух блоков (рис. 1).

Как правило, реализация этих модулей зависит друг от друга. В общем случае допуск нового потока должен осуществляться в случае, если выполняется следующее условие:

$$P \leq \sum_{i=1}^n P_i^{\min}, \tag{1}$$

где P – полоса пропускания, обеспечиваемая узлом; P_i^{\min} – минимально гарантированная полоса для потока i .

На сумму максимально гарантированных полос P_i^{\max} условий не налагается. Задача планировщика – обеспечить такую очередность отправки пакетов, чтобы соблюсти требования QoS для каждого из потоков:

$$\begin{aligned} P_i^{\max} &\geq P_i \geq P_i^{\min}, \\ L_i &\leq L_i^{\max}, \\ J_i &\leq J_i^{\max}, \end{aligned} \tag{2}$$

где P_i – объём отправленных для потока i данных; P_i^{\max} и P_i^{\min} – ограничения максимального и минимального трафика соответственно; L_i – наибольшая задержка среди пакетов в очереди станции на отправку; L_i^{\max} – максимально допустимая задержка; J_i – джиттер: разность между максимальной и минимальной L ; J_i^{\max} – максимально допустимая величина джиттера.

Потери полосы из-за фрагментации пакетов

Однако даже выполнение условия (1) не гарантирует возможности работы системы без нарушения QoS из-за наличия фрагментации пакетов. Пакет не может быть распакован до тех пор, пока у принимающей стороны не окажутся все его фрагменты. Если в (2) наблюдается равенство, то при наличии фрагментации и полной востребованности минимальной полосы возникает нарушение QoS (рис. 2).

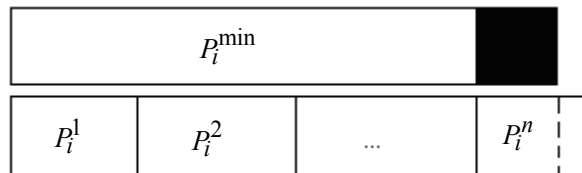


Рис. 2. Фрагментация пакета относительно минимальной полосы

Фрагмент минимальной гарантированной полосы, закрашенный чёрным, фактически потерян в текущем кадре, и минимальная полоса не соблюдается. Очевидно, что объём потерь полосы зависит от величины кадра. Если предположить, что объём пакетов, передаваемых по сети, подчиняется нормальному закону распределения, то можно попробовать оценить среднюю величину потерь от фрагментации пакетов. В ходе эксперимента полоса многократно заполнялась пакетами, размеры которых определялись на основе закона нормального распределения, фиксировался объём фрагмента, не поместившийся в полосу.

$$\Delta = P_i^{\min} - \sum_{i=1}^N Pkg_i, \tag{3}$$

где Δ – фрагмент пакета, превысивший размерность P_i^{\min} , P_i^{\max} – минимальная гарантированная полоса потока; Pkg_i – пакет с размером, полученным на основе нормального распределения.

При этом выбирается наибольшее N , при котором величина Δ положительна. Исходные данные для эксперимента приведены в табл. 2.

Таблица 2

Исходные данные для эксперимента	
Величина	Значение
Средний размер пакета	50
σ размера пакета	20
P_{\min_i} (минимальная полоса пропускания потока)	500
Количество итераций заполнения очереди	10000

Результаты эксперимента сведены в гистограмму (рис. 3).



Рис. 3. Распределение фрагментов пакетов, не помещившихся в полосу

Полученную картину следует проверить на нормальность распределения. Сама по себе сумма нормальных случайных величин есть нормальная случайная величина со средним, являющимся суммой средних, и со среднеквадратическим отклонением (СКО), являющимся суммой СКО. Далее, следуя правилу трех сигм, можно минимизировать вероятность потери полосы за счёт фрагментации. Следует отметить, что данная картина не всецело характеризует вероятность нарушения QoS вследствие недостатка полосы. На основании этих данных возможно выработать рекомендации для реализации модуля доступа к среде. Для полноты требуется рассмотреть ещё и вероятность наличия фрагментов, оставшихся от предыдущего кадра. Полученный же результат характеризует худший вариант, когда фрагментов пакетов на начало текущего кадра не имеется. Данный вопрос подробнее будет рассмотрен в последующих работах.

Алгоритмы планирования

Ряд авторов (в частности [7]) в качестве алгоритмов для планирования QoS предлагают использовать связку из DFPQ для планирования среди классов потоков (см. табл. 1) и набор из EDF, WFQ, BE или RR для планирования в самих классах. Потоки типа UGS и ERT получают место в полосе на первом этапе планирования. Поток BE не налагает требований к каналу связи, поэтому потоки этого типа получают ту часть полосы, которая останется от планирования других типов потоков. Основными объектами работы алгоритмов являются потоки типов RT и NRT. DFPQ отдаёт зарезервированное, но не занятое потоком более высокого класса место в полосе следующему за ним по приоритету. В наихудшем случае все потоки будут обладать только минимальным зарезервированным местом для отправки.

При этом WFQ гарантирует равномерную отправку всеми потоками. То есть гарантируется QoS P_{\min_i} для NRT. К отправке выбирается наиболее дискриминированный на прошлых шагах поток. Алгоритм займёт всю полосу полностью только в идеальном случае, когда будет доступна его реализация в качестве побитового конвейера. В практической его реализации следует учитывать дискрет этого конвейера (чаще всего выбирают пакет). Для этого целесообразно использовать методику из предыдущего пункта. В дополнение к традиционной реализации, приведённой в [7], следует добавить механизм контроля за соблюдением верхней границы P_i^{\max} , не учитывая при планировании те потоки, которые исчерпали свою полосу P_{\max_i} в текущей секунде.

Алгоритм EDF в его классическом виде также не учитывает один существенный момент: не фиксирует превышение потоком его P_i^{\min} . В случае, если один из потоков с небольшим L_i^{\max} произведёт вброс пакетов суммарным объёмом превышающим P_i^{\max} , то при полной загрузке полосы

произойдёт нарушение QoS для потоков (возможно нескольких) с более низким значением L_i^{\max} в этом или в последующих фреймах. Проиллюстрируем эту ситуацию (рис. 4).

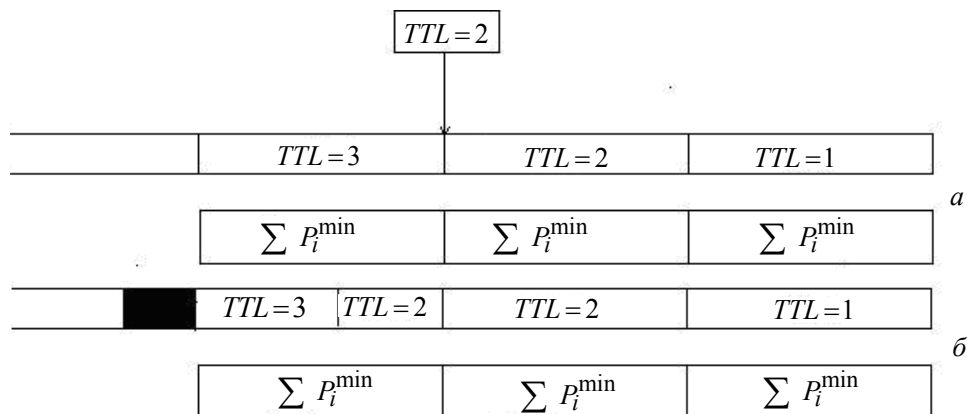


Рис. 4. Пример нарушения QoS в EDF при отсутствии контроля за P_i^{\min} потока; *a* – ситуация до вставки; *б* – после

Алгоритм гарантирует, что пакеты будут отправлены без нарушения максимального времени ожидания, если это вообще возможно для данного набора пакетов. Рассмотрим предельный случай (рис. 4, *a*), когда выполняется равенство для условия (1) и для пакетов, отправляемых в кадре N , время жизни составляет также N . То есть все потоки исчерпали свой минимум P_i^{\min} .

Допустим, одна из станций с L_i^{\max} , равным двум, производит вставку (рис. 4, *б*). В этом случае отправка всех пакетов с меньшим TTL сдвигается на более поздний срок. При этом часть пакетов из последующих кадров не сможет быть отправлена в срок (на рисунке отмечена чёрным). При этом возможны нарушения QoS сразу для нескольких потоков. То есть система обеспечения QoS должна гарантировать потоку, по крайней мере, полосу не меньшую чем P_i^{\max} с задержкой не более L_i^{\max} . Если приложение отправляет пакеты сверх этой меры, то их своевременная отправка не гарантирована, что может привести к нарушению QoS.

Предлагается ввести в EDF контроль над количеством отправляемых потоком данных таким образом, чтобы при избытке пакетов в очереди требования QoS нарушались только для одного этого потока. А именно: сохраним приоритетную очередь для пакетов по TTL, но будем на первом этапе выбирать из очереди только пакеты для тех потоков, для которых P_i^{\min} за текущую секунду ещё не был заполнен. Таким образом, мы гарантируем, что каждый из потоков получит свою минимальную полосу. Если остаётся ещё неизрасходованная полоса – производится второй проход по очереди, на этот раз без оглядки на P_i^{\min} . Однако следует для потоков RT, NRT, BE не допускать превышения P_i^{\max} . При достижении верхней границы, до истечения текущей секунды, следует считать, что очередь данного потока на отправку пуста.

Заключение

В результате проведённого исследования:

1. Была предложена методика оценки величины дополнительной полосы, требуемой для передачи фрагментированных пакетов.
2. Была произведена оценка дополнительной полосы для пакетов с заданным законом распределения размера.
3. Была предложена модификация алгоритма EDF с целью минимизации количества потоков с отказом QoS.

Исследования выполняются в рамках государственных контрактов № 13.G25.31.0011 от 07 сентября 2010 г., и № 14.740.11.0398 от 20 сентября 2010 г.

Литература

1. RTP: A Transport Protocol for Real-Time Applications [Электронный ресурс]. – Режим доступа: <http://tools.ietf.org/html/rfc3550>, свободный (дата обращения: 25.04.2012).

2. Real Time Streaming Protocol (RTSP) Applications [Электронный ресурс]. – Режим доступа: <http://tools.ietf.org/html/rfc2326>, свободный (дата обращения: 25.04.2012).
3. Resource ReSerVation Protocol (RSVP) [Электронный ресурс]. – Режим доступа: <http://www.ietf.org/rfc/rfc2205.txt>, свободный (дата обращения: 25.04.2012).
4. Integrated Services in the Internet Architecture: an Overview [Электронный ресурс]. – Режим доступа: <http://tools.ietf.org/html/rfc1633>, свободный (дата обращения: 25.04.2012).
5. An architecture for differentiated services [Электронный ресурс]. – Режим доступа: <http://tools.ietf.org/html/rfc2475>, свободный (дата обращения: 25.04.2012).
6. IEEE Standard for local and metropolitan area networks. Part 16: Air interface for broadband wireless access systems // IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004), vol., no., pp.C1-2004, May 29 2009 doi: 10.1109/IEEESTD.2009.5062485) [Электронный ресурс]. – Режим доступа: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5062485&isnumber=5062484>, для зарегистрированных пользователей (дата обращения: 25.04.11).
7. Аунг Мью Маунг. Исследование и разработка алгоритмов планирования и приоритетного управления доступом в сетях WiMax: автореф. дис. ... канд. техн. наук. – М.: МИЭТ, 2010. – 25 с.

Крупский Александр Сергеевич

Аспирант каф. автоматизированных систем управления ТУСУРа

Тел.: 8-923-420-93-52

Эл. почта: kas@asu.tusur.ru

Бойченко Иван Валентинович

Канд. техн. наук, доцент, докторант ТУСУРа

Тел.: 8-906-958-24-83

Эл. почта: biv@asu.tusur.ru

Krupskiy A.S., Boichenko I.V.

Fragmentation aware QoS scheduling algorithms in WiMax networks

The practical application of algorithms EDF and WFQ is reviewed in sphere of QoS scheduling. The differences in the practical application and in the ideal occasion of application is given. The method of grading of fragmentation overhead is represented.

Keywords: QoS, scheduling, fragmentation, overhead