

## PHB для ускоренной пересылки

### An Expedited Forwarding PHB

### Статус документа

В этом документе содержится проект стандартного протокола, предложенного сообществу Internet. Документ служит приглашением к дискуссии в целях развития и совершенствования протокола. Текущее состояние стандартизации протокола вы можете узнать из документа "Internet Official Protocol Standards" (STD 1). Документ может распространяться без ограничений.

### Авторские права

Copyright (C) The Internet Society (1999). All Rights Reserved.

### Аннотация

Определение поэтапного режима пересылки (PHB<sup>1</sup>) является важнейшей частью работы группы Diffserv. В этом документе описан PHB для ускоренной пересылки (EF<sup>2</sup>). Мы покажем общность этого PHB демонстрацией возможности его реализации с использованием разных механизмов и дадим пример использования для создания по крайней мере одного сервиса - VLL<sup>3</sup>. Приведён код, рекомендуемый для этого PHB.

Документ в формате pdf (на английском языке) доступен по ссылке [ftp://ftp.ee.lbl.gov/papers/ef\\_phb.pdf](ftp://ftp.ee.lbl.gov/papers/ef_phb.pdf)

## 1. Введение

Сетевые узлы, поддерживающие дифференцированное обслуживание для IP, используют код в заголовке IP для выбора поэтапного поведения (PHB), как специфического режима режима пересылки данного пакета [RFC2474, RFC2475]. В этом документе описан PHB для ускоренной пересылки (EF). EF PHB может использоваться для построения услуг с малыми задержками и их вариациями, а также с незначительной потерей пакетов в сквозном режиме через домены DS. Со стороны конечных точек такой сервис выглядит подобно соединению «точка-точка» или «виртуальной арендованной линии». Этот вид сервиса был также описан, как «надёжное обслуживание» (Premium service) в [2BIT]. Потери, задержки и их вариации возникают при размещении трафика в очередях на пути через сеть. Следовательно, обеспечения малых потерь, задержек и их вариаций для некоего агрегата трафика означает, что этот трафик не простаивает в очередях или находится там очень короткое время. Очереди возникают, когда скорость поступления трафика на узле (кратковременно) превышает скорость его дальнейшей отправки. Таким образом, сервис, который гарантирует отсутствие очередей для некоего агрегата, эквивалентен такому ограничению скорости, при котором на каждом транзитном узле максимальная скорость прибытия пакетов данного агрегата будет меньше минимальной скорости их дальнейшей отправки.

Создание такого сервиса включает две части:

- 1) Настройка узлов таким образом, чтобы для агрегата обеспечивалась хорошо определённая минимальная скорость отправки («хорошо определённая» скорость отправки не зависит от состояния узла и, в частности, от интенсивности другого трафика через этот узел).
- 2) Кондиционирование агрегата (с помощью политики или формовки) так, чтобы скорость прибытия на любой узел всегда была меньше заданной для узла минимальной скорости отправки.

EF PHB обеспечивает первую часть сервиса. Для обеспечения второй части служат кондиционеры трафика на границе сети, описанные в [RFC2475].

EF PHB не является обязательной компонентой архитектуры дифференцированного обслуживания, т. е., узел не обязан реализовать EF PHB для того, чтобы быть совместимым с DS. Однако, когда поддерживающий DS узел заявляет реализацию EF PHB, эта реализации должна соответствовать содержащейся здесь спецификации.

В следующих параграфах приведено подробное описание EF PHB и даны примеры возможных реализаций. Ключевые слова **необходимо** (MUST), **недопустимо** (MUST NOT), **требуется** (REQUIRED), **следует** (SHOULD), **не следует** (SHOULD NOT), **возможно** (MAY), в данном документе должны интерпретироваться в соответствии с [Bradner97].

## 2. Описание поэтапного режима ускоренной пересылки

EF PHB определяется, как трактовка пересылки для отдельного агрегата diffserv<sup>4</sup>, при которой скорость отправки пакетов данного агрегата с любого узла diffserv не меньше заданного значения. **Следует** обеспечивать гарантию

<sup>1</sup>Per-hop forwarding behavior - поэтапный режим пересылки.

<sup>2</sup>Expedited Forwarding.

<sup>3</sup>Virtual Leased Line - виртуальная арендованная (выделенная) линия.

<sup>4</sup>Differentiated Service - дифференцированное обслуживание. *Прим. перев.*

заданной скорости отправки независимо от интенсивности остального трафика, пытающегося проходить через узел. **Следует** обеспечивать среднее значение скорости<sup>1</sup> отправки не менее заданного значения (поведение в течение более коротких интервалов не задаётся осознанно). Заданное минимальное значение скорости **должно** быть настраиваемым сетевым администратором (с использованием любого, поддерживаемого узлом, механизма).

Если EF PNB реализуется с помощью механизма, который разрешает неограниченные преимущества перед другим трафиком (например, очереди с разным приоритетом), реализация **должна** включать те или иные меры по ограничению негативного влияния трафика EF на другой трафик (например, ограничение скорости с помощью метода token bucket - переполняющаяся ёмкость). Трафик, превышающий заданный предел **должен** отбрасываться. Максимальная скорость EF и величина пиков (если это применимо) **должны** быть настраиваемыми сетевым администратором (с использованием любого, поддерживаемого узлом, механизма). Минимальная и максимальная скорость могут совпадать и задаваться в этом случае одним параметром.

В Приложении описано возможное применение данного PNB для организации сквозного сервиса.

## 2.2 Примеры механизмов для реализации EF PNB

Несколько механизмов планирования очередей могут использоваться для обеспечения пересылки пакетов в соответствии с описанием параграфа 2.1 и, соответственно, реализации EF PNB. Простая очередь с приоритизацией будет обеспечивать требуемое поведение, если нет очереди с более высоким приоритетом, способной принять трафик EF со скоростью, превышающей заданную (это можно обеспечить с помощью политики контроля скорости типа token bucket, связанной с каждой очередью, принимать другой трафик).

Можно также использовать одну очередь из группы очередей, обслуживаемых планировщиком со взвешенным круговым обходом, где часть выходной полосы, отдаваемая для очереди EF, эквивалентна заданной скорости. Это можно реализовать, к примеру, с использованием одного PNB или совместимого с селекторами классов набора PNB [RFC2474].

Другой возможной реализацией является планировщик CBQ [CBQ], который отдаёт очереди EF приоритет вплоть до заданной скорости.

Все эти механизмы обеспечивают базовые свойства, требуемые для реализации EF PNB, хотя каждый из вариантов будет иметь дополнительные особенности поведения типа вариаций задержки для отдельных микропотоков. В приложении A.3 рассматриваются некоторые из таких различий.

## 2.3 Рекомендуемый код для данного PNB

Для EF PNB рекомендуется код 101110.

## 2.4 Изменяемость

Пакеты, маркированные для EF PNB, **могут** быть перемаркированы на границе домена DS только с использованием других кодов, удовлетворяющих EF PNB. Пакеты, маркированные для EF PNB, домену DS **не следует** перемещать (поднимать или опускать) в другие PNB.

## 2.5 Туннелирование

При туннелировании пакетов EF туннелирующие пакеты должны маркироваться, как EF.

## 2.6 Взаимодействие с другими PNB

При выполнении требований параграфа 2.1 на узле или в домене DS могут развёртываться другие PNB или группы PNB.

## 3. Вопросы безопасности

Для защиты себя от атак на отказ служб (DoS<sup>2</sup>) периметр домена DS **должен** строго проверять для всех пакетов с маркировкой EF соответствие скорости согласованному со смежным восходящим доменом значению (скорость должна быть не больше значения, заданного для EF PNB). Пакеты, вызывающие превышение согласованной скорости **должны** отбрасываться. Если смежные домены не согласовали скорость EF, нисходящий домен **должен** использовать нулевое значение скорости (т. е., отбрасывать все пакеты с маркировкой EF)<sup>3</sup>.

Поскольку сквозное приоритетное обслуживание на базе EF PNB требует, чтобы восходящий домен применял политику и формовал трафик с маркировкой EF для обеспечения скорости, согласованной с нисходящим доменом, средствам реализации политики в нисходящем домене никогда не следует отбрасывать пакетов. Поэтому факты такого отбрасывания **следует** отмечать (например, с помощью прерываний SNMP), как потенциальное нарушение безопасности или серьёзные ошибки в конфигурации. Аналогично, в результате ограничения скорости трафика агрегата EF на каждом узле, очередь EF в любом случае не должна переполняться и при возникновении фактов отбрасывания пакетов, такие события **следует** отмечать, как возможную атаку или серьёзную конфигурационную ошибку.

## 4. Взаимодействие с IANA

В этом документе выделяется один код 101110 из пула 1 (Pool 1) пространства кодов, определённого в [RFC2474].

## 5. Литература

[Bradner97] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, [RFC 2119](#), March 1997.

<sup>1</sup>При измерении за период не менее, чем потребуется для передачи в канал с такой скоростью пакетов размера MTU.

<sup>2</sup>Denial of service attack.

<sup>3</sup>См. обсуждение в разделе 6 [RFC 3260](#). Прим. перев.

- [RFC2474] Nichols, K., Blake, S., Baker, F. and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", [RFC 2474](#), December 1998.
- [RFC2475] Black, D., Blake, S., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services", [RFC 2475](#), December 1998.
- [2BIT] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet", Work in Progress<sup>1</sup>, <ftp://ftp.ee.lbl.gov/papers/dsarch.pdf>
- [CBQ] S. Floyd and V. Jacobson, "Link-sharing and Resource Management Models for Packet Networks", IEEE/ACM Transactions on Networking, Vol. 3 no. 4, pp. 365-386, August 1995.
- [RFC2415] Poduri, K. and K. Nichols, "Simulation Studies of Increased Initial TCP Window Size", RFC 2415, September 1998.
- [LCN] K. Nichols, "Improving Network Simulation with Feedback", Proceedings of LCN '98, October 1998.

## 6. Адреса авторов

### Van Jacobson

Cisco Systems, Inc  
170 W. Tasman Drive  
San Jose, CA 95134-1706  
E-Mail: [van@cisco.com](mailto:van@cisco.com)

### Kathleen Nichols

Cisco Systems, Inc  
170 W. Tasman Drive  
San Jose, CA 95134-1706  
E-Mail: [kmn@cisco.com](mailto:kmn@cisco.com)

### Kedarnath Poduri

Bay Networks, Inc.  
4401 Great America Parkway  
Santa Clara, CA 95052-8185  
E-Mail: [kpoduri@baynetworks.com](mailto:kpoduri@baynetworks.com)

## Перевод на русский язык

Николай Малых  
[nmalykh@protokols.ru](mailto:nmalykh@protokols.ru)

## Приложение А: Пример использования и опыт работы с EF PHB

### А.1 Услуги VLL

Услуги VLL, называемые также приоритетным обслуживанием (Premium service) [2BIT] характеризуются пиковой полосой пропускания.

### А.2 Опыт использования в ESNET

Прототип услуг VLL был развернут в опорной сети DOE ESNet с использованием очередей со взвешенным круговым обходом на базе маршрутизаторов Cisco серии 75xx для реализации EF PHB. Первые тесты были весьма успешными и продолжается работа по выводу сервиса на коммерческую эксплуатацию (см. <ftp://ftp.ee.lbl.gov/talks/vj-doeqos.pdf> и <ftp://ftp.ee.lbl.gov/talks/vj-i2qos-may98.pdf>).

### А.3 Результаты моделирования

#### А.3.1 Вариации задержки

В параграфе 2.2 было отмечено, что для реализации EF PHB может использоваться множество механизмов. Простейшим вариантом является PQ<sup>2</sup>, где скорость поступления пакетов в очередь строго меньше скорости услуги<sup>3</sup>. Поскольку задержки в очередях на пути доставки приводят к вариациям скорости, особенностью данного варианта является то, что микропотоки с маркировкой EF будут сталкиваться с незначительными вариациями скорости обслуживания. EF PHB не содержит явных требований к вариациям, но из определения ясно, что ожидаемые вариации для потока пакетов сервиса на основе EF PHB при использовании PQ будут меньше, нежели при обычной доставке<sup>4</sup>.

<sup>1</sup>Работа завершена в RFC 2638. *Прим. перев.*

<sup>2</sup>Priority queue - очередь с приоритетом.

<sup>3</sup>Subscribed rate.

<sup>4</sup>Best-effort delivery - доставка по мере возможности.

Были проведены эксперименты на макете для сравнения взвешенных очередей с циклической обработкой (WRR<sup>1</sup>) и PQ в плане вариаций. Эти два варианта были выбраны потому, что они представляют худший и лучший<sup>2</sup> случай, соответственно, в плане возникновения вариаций и мы хотели предоставить разработчикам EF некие рекомендации по использованию WRR или аналогичных механизмов.

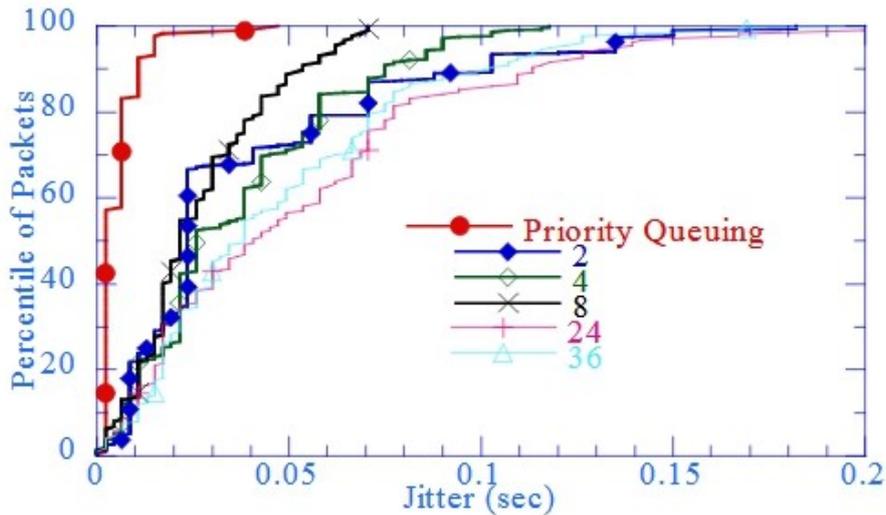


Рисунок 1.

Экспериментальная сеть представляла собой модифицированный вариант стенда ns-2, описанного в [RFC2415] и [LCN]. Модули CBQ из состава ns-2 использовались в качестве базы для реализации приоритетной очереди м WRR. Топология модели включала шесть интервалов маршрутизации (hop) с уменьшающейся пропускной способностью в направлении одного узкого канала 1,5 Мбит/с (см. рисунок 6). Отправители передают пакеты с маркировкой EF со средней скоростью, равной скорости услуги для них. Пакеты генерируются с разбросом ±10% в межпакетных интервалах относительно номинального значения для скорости услуги. Потоки отдельных источников объединялись в агрегат со скоростью 30% от полосы самого узкого канала (450 кбит/с). Для заполнения канала использовалась смесь пакетов FTP и HTTP. Каждый из источников пакетов EF генерировал пакеты одного размера 160 или 1500 байтов.

Хотя статистика представлена для потоков с одним размером пакетов, во всех экспериментах использовалась смесь коротких и длинных пакетов EF, поэтому очереди EF включали пакетов обоих размеров.

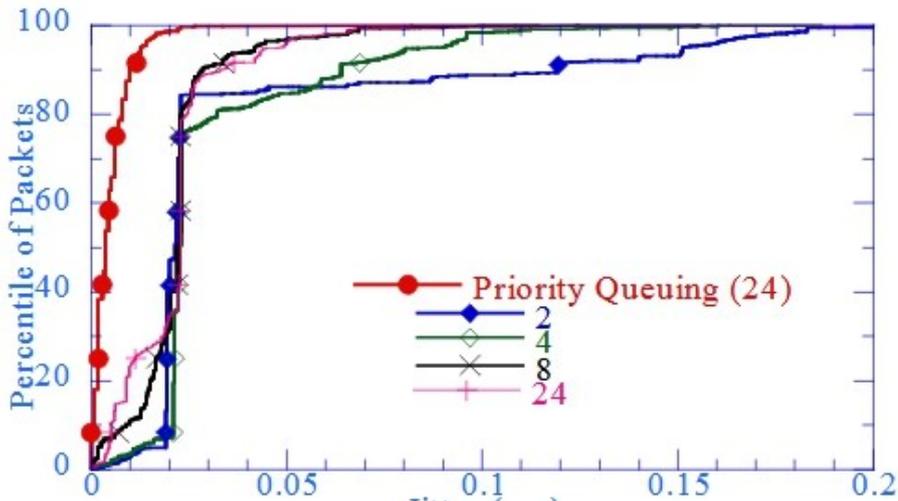


Рисунок 2.

Определим вариации (jitter) как абсолютное значение разности между интервалами от поступления до отправки двух смежных пакетов  $(|(a_j - d_j) - (a_i - d_i)|)^3$ . Для целевого потока в каждом эксперименте записывались значения вариаций для 50 и 90% пакетов, выраженные в процентах от скорости услуги EF, представленные в таблицах. Документ в формате pdf содержит также графики для вариаций.

Таблица 1. Зависимость вариаций от числа потоков EF.

	1500 байтов		160 байтов	
	50%	90%	50%	90%
PQ (24)	1	5	17	43
2	11	47	96	513
4	12	35	100	278
8	10	25	96	126
24	18	47	96	143

В экспериментах сравнивались вариации для реализаций EF PHB на основе WRR и PQ. Оценивалось влияние различного выбора веса очередей WRR и числа очередей на вариации. Для WRR определено отношение скоростей обработка/прибытие как отношение скорости обслуживания для очереди EF (или минимальный расход очереди выходного канала) к пиковой скорости прибытия в очередь пакетов с маркировкой EF. Результат не будет стабильным, если вес WRR выбран для точного балансирования скоростей прибытия и отправки, поэтому используется минимальное отношение обслуживание/прибытие 1,03. В проведённых экспериментах это означает, что очередь EF

<sup>1</sup>Weighted round-robin.

<sup>2</sup>В оригинале ошибочно указано наоборот. Прим. перев.

<sup>3</sup>a – время прибытия, d – время отправки пакетов. Прим. перев.

получает не менее 31% выходных каналов. В экспериментах с WRR канал заполнялся до конца другим трафиком, как описано выше, с разделением трафика без маркировки EF между очередями, не относящимися к EF (из описания эксперимента должно быть ясно, что была предпринята попытка внести наибольшие вариации и не делается предположений, что такие установки и трафик отражают нормальную ситуацию).

Таблица 2. Зависимость вариаций от отношения скорости обработки и прибытия.

WRR	1500 байтов		160 байтов	
	50%	90%	50%	90%
Обработка/прибытие				
PQ	1	3	17	43
1.03	14	27	100	178
1.30	7	21	65	113
1.50	5	13	57	104
1.70	5	13	57	100
2.00	5	13	57	104
3.00	5	13	57	100

В первой серии экспериментов использовалось минимальное отношение скоростей обслуживания/прибытия 1,06 и менялось число отдельных микропотоков, составляющих агрегат EF, от 2 до 36. Проведено сравнение полученных данных с результатами для PQ при 24 потоках. Сначала проверялся микропоток 1500-байтовых пакетов, передаваемых со скоростью услуги 56 кбит/с, а затем эксперимент повторялся для пакетов размером 160 байтов. Таблица 2 показывает значения вариаций для 50 и 90% пакетов (в процентах от времени передачи пакета при скорости услуги). Рисунок 1 показывает график для потока пакетов размером 1500 байтов, а рисунок 2 - для пакетов размером 160 байтов. Отметим, что «время пакета» размером 1500 байтов при скорости 56 кбит/с составляет 214 мсек, а для пакета в 160 байтов - 23 мсек. Вариации для больших пакетов редко превышают половину времени передачи пакета, хотя для мелких пакетов в большинстве случаев вариации не меньше времени передачи одного пакета. Следует принимать во внимание, что агрегат EF представляет собой смесь больших и мелких пакетов, поэтому короткие пакеты могут ожидать обработки длинных пакетов в очереди EF. PQ обеспечивает существенно меньшие вариации.

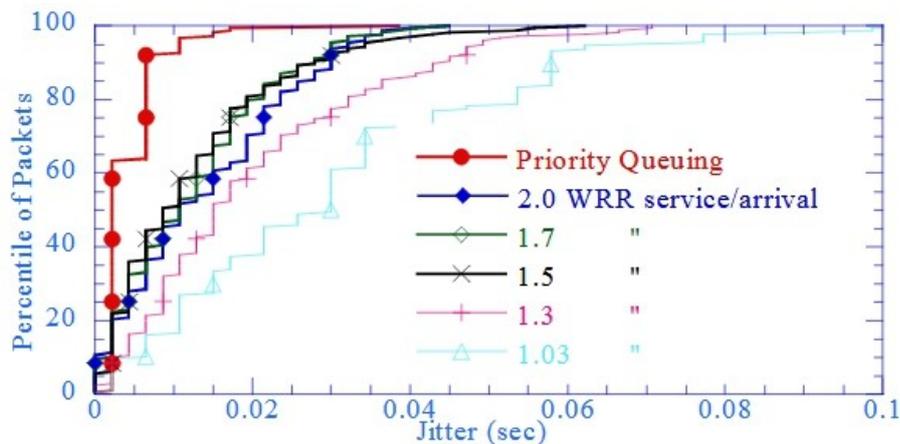


Рисунок 3. Вариации задержки для 8 потоков EF с пакетами по 1500 байтов.

Далее исследовалось влияние роста отношения скоростей обслуживания/прибытия. Такой рост означает сокращение времени пребывания пакетов EF в очереди при сохранении полосы, доступной для других очередей. В этих экспериментах число потоков в агрегате EF было равно восьми, а общее число очередей - 5 (4 очереди для потоков без EF). Таблица 3 показывает результаты для потоков с размером пакетов 1500 и 160 байтов. Рисунок 3 показывает графики для пакетов размером 1500 байтов, а рисунок 4 для пакетов размером 160 байтов. Повышение производительности наблюдается до отношения обработка/прибытие равного 1,5. Отметим, что большие значения

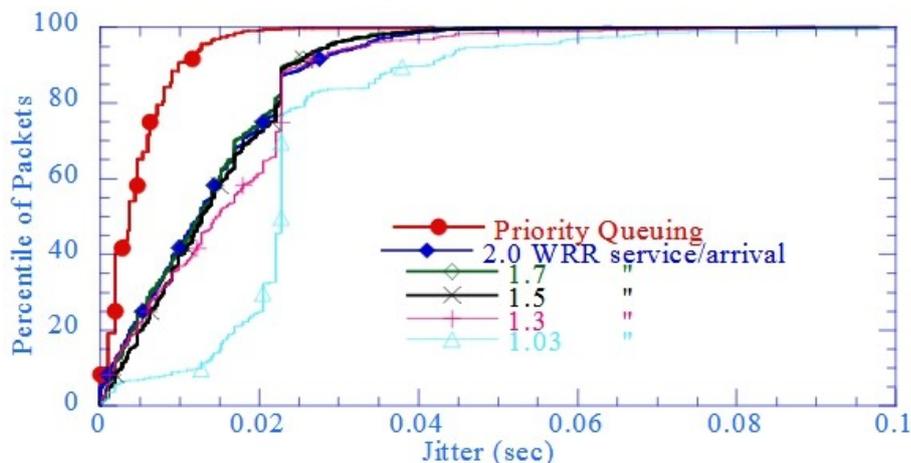


Рисунок 4. Вариации задержки для 8 потоков EF с пакетами по 160 байтов.

отношения скоростей обработки/прибытия не дают такой же производительности, как PQ, но в этом случае для 90% пакетов вариации не превышают время доставки даже для мелких пакетов.

Таблица 3. Зависимость вариаций от числа очередей на выходном интерфейсе.

Число потоков EF	1500 байтов	
	50%	90%
PQ (8)	1	3
2	7	21
4	7	21
6	8	22
8	10	23

Увеличение числа очередей на выходных интерфейсах может привести к росту вариаций времени обслуживания для пакетов EF, поэтому были проведены эксперименты с разным числом очередей на каждом выходном порту. Число потоков в агрегате было равно 8 и использовалось минимальное отношение скоростей обработки/прибытия 1,03. Результаты показаны на рисунке 5 и в таблице 3. Рисунок 5 включает также график для PQ с 8 потоками.

Видно, что в большинстве случаев вариации для WRR достаточно малы и могут быть дополнительно снижены выбором корректного соотношения между полосой очереди WRR на выходном канале и скоростью сервиса EF. Как было отмечено, WRR представляет собой худший случай, а PQ - лучший. Другие возможности включают WFQ или CBQ с фиксированным пределом скорости для очереди EF и предоставлением этой очереди приоритета по сравнению с другими очередями. Мы предполагаем, что это обеспечит производительность, близкую к PQ, хотя для проверки этого утверждения потребуются дополнительные эксперименты. Мы пока не исследовали систематически влияние числа интервалов маршрутизации, выделения для EF другой (не 30%) доли полосы канала и более сложной топологии. Приведённая здесь информация не является частью спецификации EF PHB и дана просто в качестве рекомендаций для разработчиков.

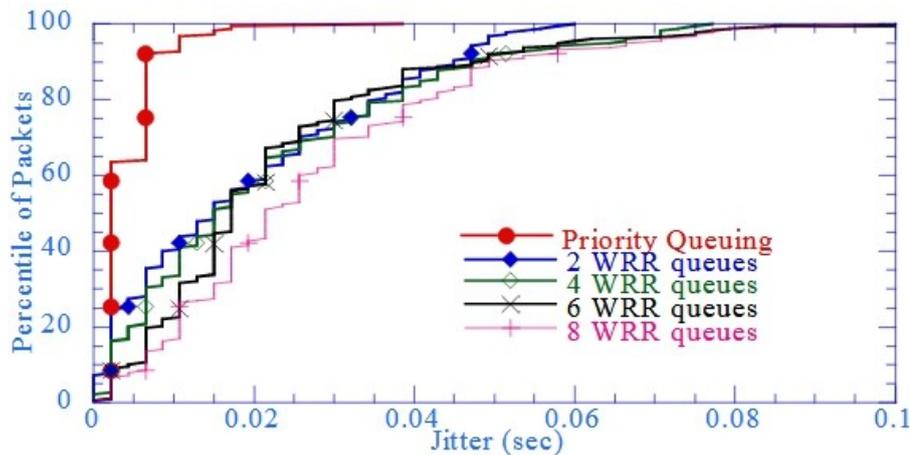


Рисунок 5. Вариации задержки при множестве очередей (8 потоков EF с пакетами по 1500 байтов).

### A.3.2 Сервис VLL

Таблица 4. Производительность FTP с использованием сервиса VLL.

Доля канала для EF (%)	Средняя скорость доставки (кбит/с)		
	Услуга	Арендованная линия VLL	
20	100	90	90
40	150	143	143
60	225	213	215

Мы использовали имитацию для проверки организации сервиса VLL на основе EF PHB. Проверялось сходство такого сервиса с арендованной линией, имеющей идентичную скорость. В этом эксперименте ни один пакет EF не был отброшен в сети и всегда достигалась нужная скорость для источников CBR<sup>1</sup>. Однако мы хотели увидеть, будет ли VLL вести себя «подобно проводам» для использующего сервис трафика TCP. Поэтому использовались долговременные соединения FTP на базе сервиса VLL. Ошибка: источник перекрёстной ссылки не найден показывает долю каждого канала, отданную для трафика EF (на каналах с меньшим числом микропотоков EF полоса была уже), скорость услуги VLL, средняя скорость для однотипной пары отправитель-получатель, подключённой к полнодуплексному выделенному каналу со скоростью услуги и средняя скорость потоков VLL для каждого теста. Потери пакетов происходили только во входном буфере формовщика, но не в сети. Максимальная скорость не была достигнута в силу особенностей протокола TCP.

<sup>1</sup>Committed Bit Rate – согласованная скорость. Прим. перев.

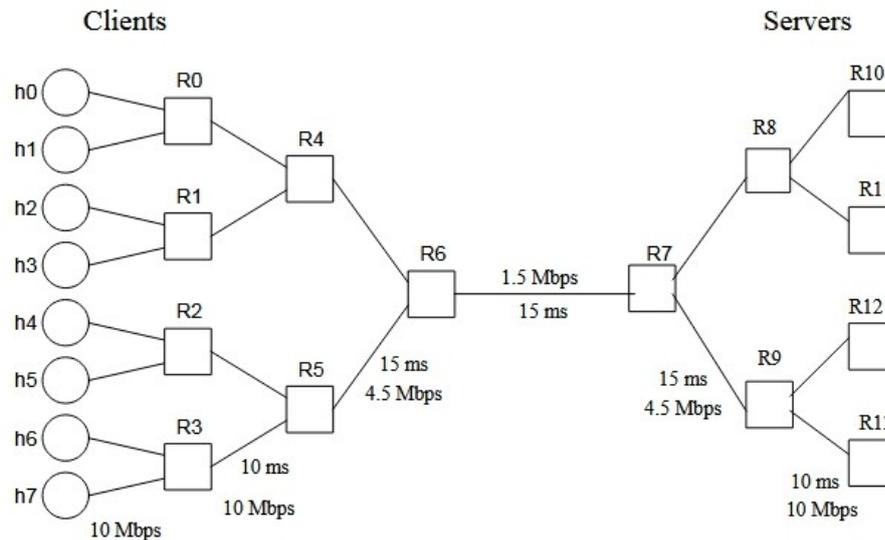


Рисунок 6. Топология модели.

## Полное заявление авторских прав

Copyright (C) The Internet Society (1999). Все права защищены.

Этот документ и его переводы могут копироваться и предоставляться другим лицам, а производные работы, комментирующие или иначе разъясняющие документ или помогающие в его реализации, могут подготавливаться, копироваться, публиковаться и распространяться целиком или частично без каких-либо ограничений при условии сохранения указанного выше уведомления об авторских правах и этого параграфа в копии или производной работе. Однако сам документ не может быть изменён каким-либо способом, таким как удаление уведомления об авторских правах или ссылок на Internet Society или иные организации Internet, за исключением случаев, когда это необходимо для разработки стандартов Internet (в этом случае нужно следовать процедурам для авторских прав, заданных процессом Internet Standards), а также при переводе документа на другие языки.

Предоставленные выше ограниченные права являются бессрочными и не могут быть отозваны Internet Society или правопреемниками.

Этот документ и содержащаяся в нем информация представлены "как есть" и автор, организация, которую он/она представляет или которая выступает спонсором (если таковой имеется), Internet Society и IETF отказываются от каких-либо гарантий (явных или подразумеваемых), включая (но не ограничиваясь) любые гарантии того, что использование представленной здесь информации не будет нарушать чьих-либо прав, и любые предполагаемые гарантии коммерческого использования или применимости для тех или иных задач.

### Подтверждение

Финансирование функций RFC Editor обеспечивается Internet Society.