Title

Настройка политик QoS

Успешно сдайте бесплатный сертификационный экзамен в Академии "Инфинет" и получите статус сертифицированного инженера Инфинет.

Пройти сертификационный экзамен

Содержание

- Введение
- Терминология
- Схема распространения пакетов
- Метрики качества
  - Потери
    - Пропускная способность
    - Пакетная производительность
  - Задержка
  - Джиттер
  - Требования сервисов к метрикам качества
- Методы обеспечения QoS
- Механизмы приоритизации трафика
  - Приоритизация в Ethernet (802.1p)
  - Приоритизация в IP
  - Установка приоритета
  - Реализация очередей в устройствах Инфинет
    - Таблица внутренней организации очередей сообщений
    - Таблица соответствия протокольных и внутренних приоритетов для устройств семейств InfiLINK 2x2, InfiMAN 2x2, InfiLINK Evolution, InfiMAN Evolution
    - Таблица соответствия протокольных и внутренних приоритетов для устройств семейств InfiLINK XG, InfiLINK XG 1000, Vector 5, Vector 6, Vector 70
  - Диспетчеризация очередей
    - Строгая диспетчеризация
    - Взвешенная диспетчеризация
  - Рекомендации по приоритизации трафика
- Механизмы ограничения пропускной способности
  - Принцип ограничения скорости в устройствах Инфинет
  - Алгоритм Token Bucket
  - Виды ограничений скорости в устройствах Инфинет
    - Таблица функциональных возможностей по ограничению пропускной способности в устройствах Инфинет
  - Рекомендации по конфигурации ограничения пропускной способности
- Дополнительные материалы
  - White papers
  - Вебинары
  - Видео
  - Прочее

Введение

Развитие сетей передачи данных влечёт за собой рост объёма передаваемого трафика, что требует использования политики качества обслуживания. Внедрение политики позволит классифицировать сетевой трафик и распределять сетевые ресурсы между классами трафика.

Терминология

- **QoS** (Quality of Service - качество обслуживания) - технология, позволяющая выполнить классификацию потока данных и приоритизировать передачу каждого потока в соответствии с его классом.
- **Политика QoS** - документ, описывающий принципы классификации потоков трафика и требования к ресурсам для каждого из классов.
- **Поток трафика** - данные одного сервиса, передающиеся между двумя узлами.
- **Сервис** - служба, запущенная на конечных узлах, выполняющие обмен данными между узлами. Данные, относящиеся к одному сервису, отличаются уникальным набором значений служебных полей и структурой сетевых пакетов. Примерами сервисов являются IP-телефония, web и видеонаблюдение.
- **Зона ответственности** - сегмент сети, за эффективное функционирование которого отвечает определённый субъект. В качестве субъекта может выступать как конкретный человек, так и организация.

- **DS-домен** (Differentiated Services домен - домен диффиренцированных сервисов) - логическая область, в которой применяются единые правила классификации трафика, определяемые политикой QoS. Обычно DS-домен совпадает с зоной ответственности.
- **CIR** (Committed Information Rate) - гарантированная пропускная способность. Система гарантирует выделение ресурсов для соблюдения CIR для сервиса.
- **MIR** (Maximum Information Rate) - максимальная пропускная способность. В случае выполнения CIR, сервисам могут быть предоставлены дополнительные ресурсы. Дополнительные ресурсы не могут превысить порог MIR и их выделение не гарантировано.

Схема распространения пакетов

В пакетных сетях передачи данных трафик распространяется от узла-отправителя к узлу-получателю через каналы связи и промежуточные устройства. В общем случае пакет данных обрабатывается каждым из промежуточных устройств независимо. Рассмотрим пример обработки пакета данных промежуточным сетевым устройством (рис. 1):

1. Узел-2 формирует пакет данных и передаёт его в Среду-2. Пакет данных инкапсулируется в кадр канального протокола, используемого в Среде-2.
2. Кадр данных распространяется в Среде-2. Для этого кадр преобразовывается в модулированный сигнал, соответствующий физическим свойствам среды. Сигналы, используемые в проводных и беспроводных средах, будут отличаться, что повлияет на эффекты их распространения и сценарии использования.
3. Сигнал поступает на входящий интерфейс устройства, демодулируется, и полученный кадр данных проверяется на целостность: если кадр повреждён, то он отбрасывается.
4. Кадр проходит этап маршрутизации, на котором определяется его дальнейший путь следования. Кадр может быть адресован Сетевому устройству, в этом случае он передаётся на обработку внутренним процессам. Кадр может быть адресован другому узлу и, в этом случае, возможны два варианта развития событий: кадр должен быть передан далее через исходящий интерфейс, либо отброшен (если Среда-2 является общей средой, то все передаваемые сигналы будут приняты всеми устройствами, подключенными к среде. В соответствии с логикой протоколов канального уровня, если в заголовке кадра в качестве получателя указан адрес, не принадлежащий устройству, то устройство должно его отбросить).
5. Если кадр должен быть обработан и передан другому узлу, то кадр поступает в очередь сообщений. Очередь сообщений представляет собой набор буферов, в которые помещаются данные, принятые входящими интерфейсами. Число и объём буферов памяти, в которых хранится очередь сообщений, не стандартизованы и зависит от производителя оборудования. Например, в устройствах семейства InfiLINK 2x2 выделено 32 очереди, 17 из которых доступны пользователю для настройки.
6. Кадр данных проходит через очередь сообщений, в которую он был помещен, и поступает в исходящий интерфейс.
7. Поскольку очереди сообщений являются связующим звеном между наборами входящих и исходящих интерфейсов, то в устройстве должен быть выделен контроллер, который выполняет заполнение очередей входящими данными и выборку из очередей для передачи исходящим интерфейсам. Как правило, эти функции выполняет центральный процессор (ЦП). Как будет показано далее, заполнение и выборка очередей может выполняться неравномерно и зависеть от классификации потоков данных.
8. Исходящий интерфейс формирует модулированный сигнал и передаёт его в Среду-5, к которой подключен Узел-5, являющийся получателем исходного пакета данных.
9. Узел-5 принимает сигнал, демодулирует его и обрабатывает полученный кадр данных.

Следует отметить, что в большинстве современных сетевых устройств сетевые интерфейсы являются комбинированными и могут выступать как в роли входящих, так и в роли исходящих.

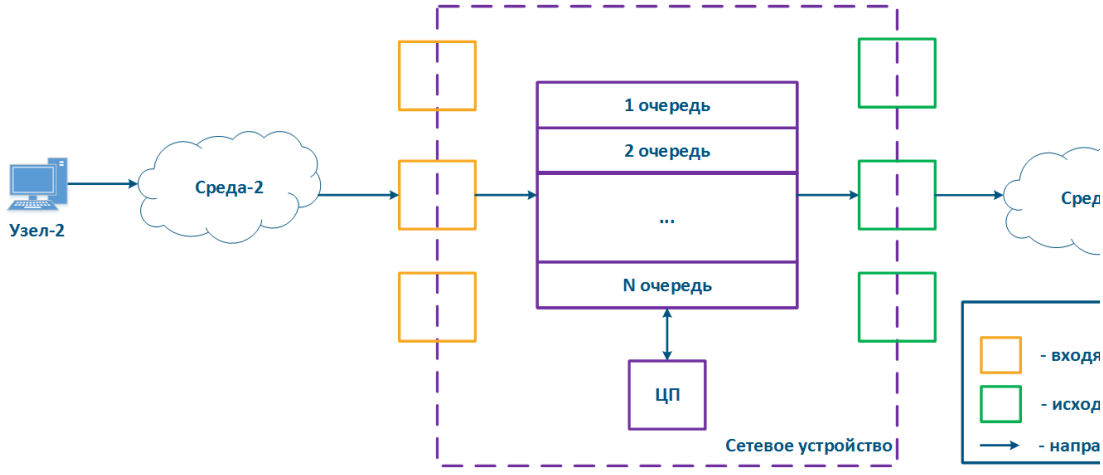


Рисунок 1 - Схема прохождения трафика через сетевое устройство

Сетевое устройство может быть промежуточным для нескольких пар узлов, каждая из которых может передавать данные нескольких сервисов (рис. 2а). Рассмотрим схему, в которой Сетевое устройство является промежуточным для трафика пар узлов Узел-1 - Узел-4, Узел-2 - Узел-5 и Узел-3 - Узел-6. Первая пара передаёт данные трёх сервисов, вторая - двух, третья - одного. В общем случае, при отсутствии настроек QoS, данные всех сервисов попадают в общую очередь в порядке поступления их на Сетевое устройство и в этом же порядке будут из очереди переданы на исходящие интерфейсы.

Title

При настроенном QoS можно классифицировать каждый из входящих потоков трафика, например, по его типу и сопоставить каждому классу отдельную очередь (рис. 2б). Каждой из очередей пакетов может быть назначен свой приоритет, который будет учитываться при извлечении пакетов из очередей сообщений, что позволит гарантировать показатели качества. Классификация потоков трафика может быть выполнена не на основании используемых сервисов, а по другим критериям. Например, каждой паре пользователей может быть выделена отдельная очередь сообщений (рис. 2в).

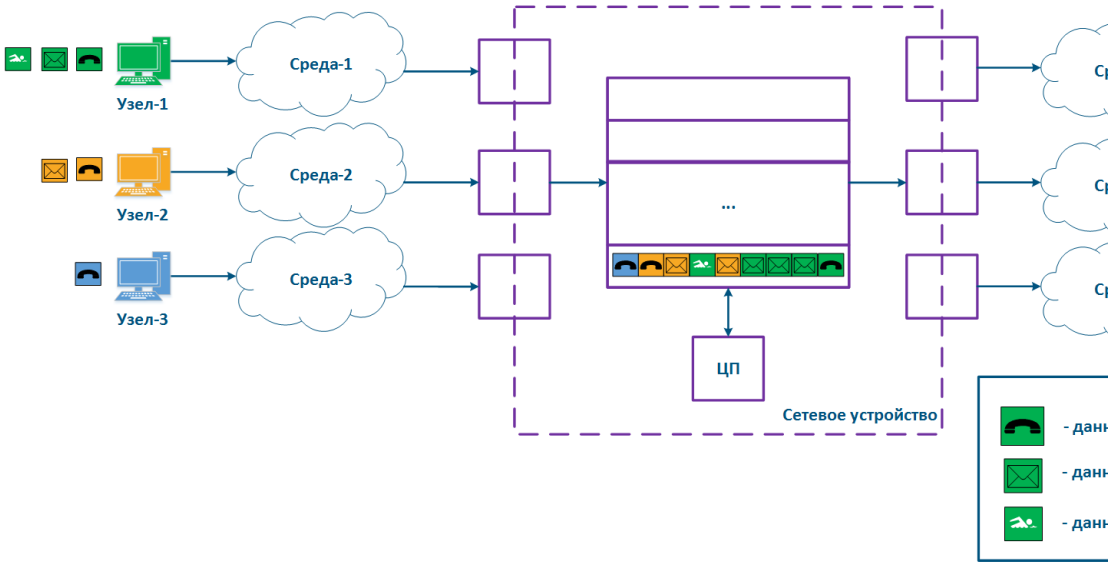


Рисунок 2а - Формирование очереди для различных сервисов без QoS

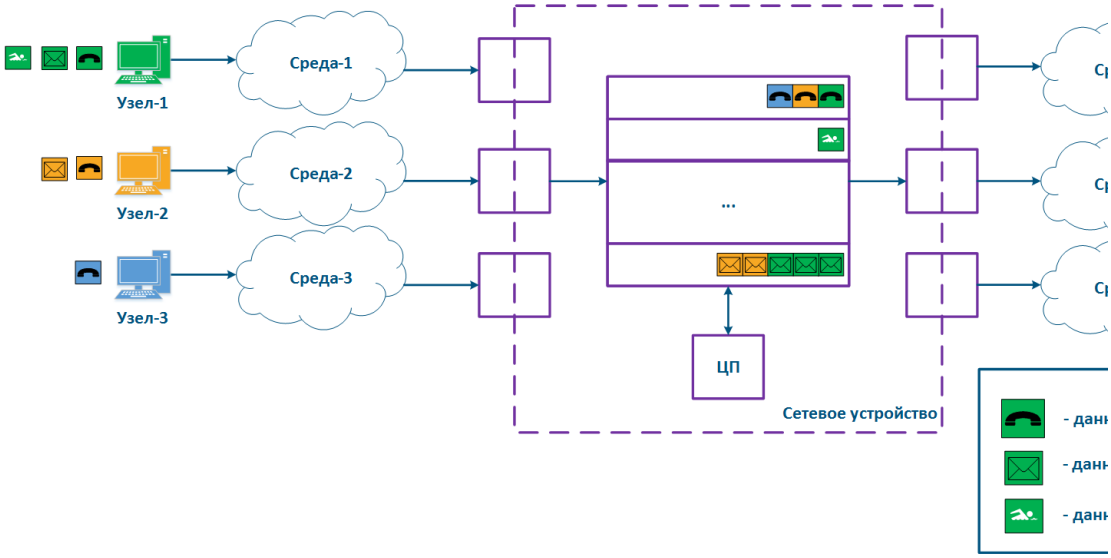


Рисунок 2б - Формирование очередей различных сервисов с QoS

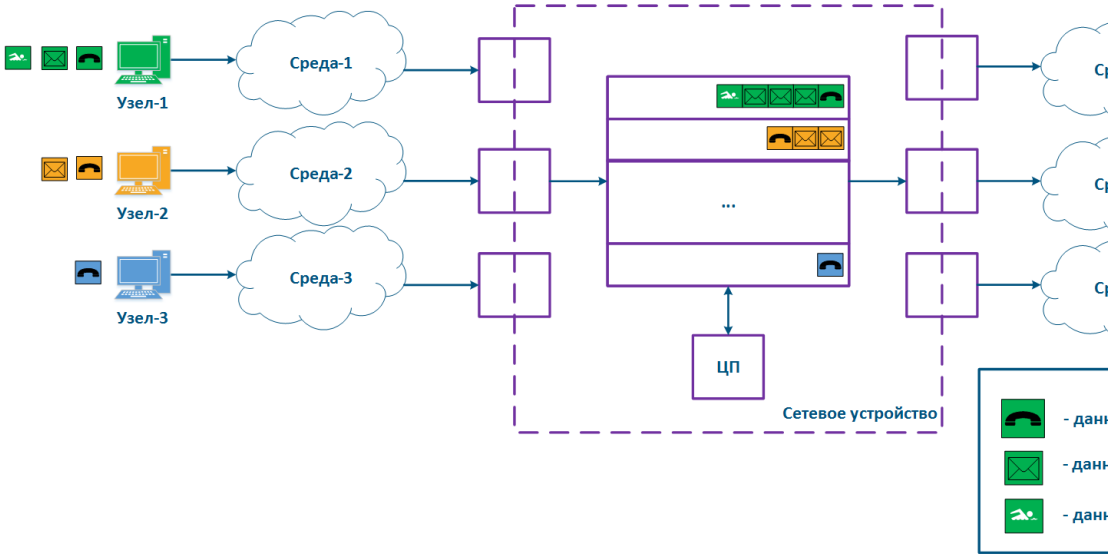


Рисунок 2в - Формирование очередей различных пользователей с QoS

Следует иметь в виду, что на пути данных от источника до получателя может быть расположено несколько промежуточных сетевых устройств, очереди сообщений на которых независимы друг от друга, т.е. эффективное внедрение политики QoS потребует конфигурации всех сетевых узлов.

Метрики качества

Основные выводы предыдущего раздела, которые будут использоваться нами для определения метрик качества:

- Пропускная способность каналов связи и сетевых устройств является конечной.
- Время доставки данных от источника к получателю является ненулевым.
- Канал связи представляет из себя среду с набором физических параметров, которые определяют эффекты при распространении сигнала.
- Программная и аппаратная архитектура сетевого устройства могут оказывать влияние на распространение данных.

Выделяют три основные метрики качества:

- Потери.
- Задержка.
- Джиттер.

Подробно рассмотрим метрики на примере: Узел-2 передаёт три пакета данных Узлу-5, источник и получателя данных соединяет промежуточное Сетевое устройство, пакеты передаются в рамках одного сервиса, т.е. их ключевые служебные поля совпадают.

Потери

При передаче потока данных, часть из них могут быть не приняты, либо приняты с ошибками. В этом случае можно говорить о потери данных, которые измеряются как отношение количества принятых пакетов к количеству переданных. В примере (рис. 3) Узел-2 передаёт пакеты с идентификаторами 1,2 и 3, однако Узел-5 принимает только пакеты 1 и 3, т.е. пакет с идентификатором 2 потерян. Существуют сетевые механизмы, позволяющие выполнить повторную передачу потерянных данных. Например, к таким механизмам можно отнести протоколы TCP и ARQ.

Причины потери данных можно выделить в следующие группы:

- **Потери в среде:** потери, связанные с распространением сигнала в физической среде. Например, кадр будет потерян, если уровень полезного сигнала ниже чувствительности приёмника. Также потери могут быть вызваны физическим повреждением интерфейсов подключения к среде или импульсные наводки, возникающие из-за некачественного заземления.
- **Потери в интерфейсе:** потери, связанные с обработкой очереди сообщений на входящем или исходящем интерфейсе. У каждого из интерфейсов существует буфер памяти, который может быть полностью заполнен при интенсивном потоке данных. В этом случае все последующие данные, поступающие в интерфейс, будут отброшены, т.к. не могут быть помещены в буфер.
- **Потери в устройстве:** данные, отброшенные в соответствии с логикой конфигурации сетевого устройства. В случае, если очереди сообщений будут переполнены, то входящие данные не смогут быть добавлены в очередь обработки, и сетевое устройство их отбросит. Также, к этим потерям можно отнести пакеты данных, отфильтрованные списками доступа и файрволом.

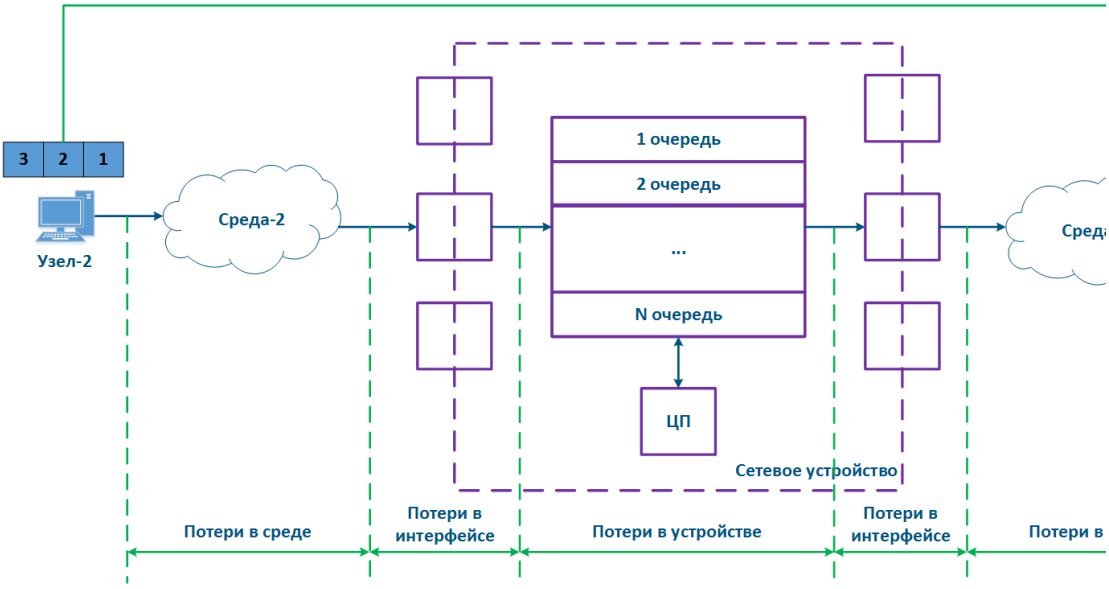


Рисунок 3 - Пример потери пакета данных

Величина потерь влияет на две метрики, которые не относят к основным: пропускная способность и пакетная производительность.

Пропускная способность

Одной из основных метрик, используемых на практике, является пропускная способность, величина которой зависит от потерь. Пропускная способность определяется физическими возможностями канала связи и возможностью обработки потока данных промежуточными сетевыми устройствами. Пропускная способность канала связи определяется как максимальный объём данных, который может быть предан от источника к получателю в единицу времени.

Пакетная производительность

Параметром, влияющим на пропускную способность и состояние очередей сообщений является пакетная производительность устройства. Под пакетной производительностью понимается максимальное число пакетов данных заданной длины, которое устройство способно передать в единицу времени.

Пропускная способность, получаемая на практике, зависит как от пакетной производительности, так и от характеристик интерфейса, поэтому на этапе про ектирования сети следует обращать внимание на согласованность этих параметров, чтобы ни одно из них не стало "бутылочным горлышком" канала связи и сетевого сегмента.

Величина пакетной производительности определяется аппаратными возможностями центрального процессора и объёмом внутренней памяти. Сетевые устройства обрабатывают множество потоков трафика с разными размерами кадров канального уровня, поэтому при тестировании производительности используются следующие значения размера кадра Ethernet:

- минимальный размер = 64 байта;
- средний размер = 512 байт;
- максимальный размер = 1518 байт.

Из-за ограниченного объёма внутренней памяти, лучшая пакетная производительность достигается для минимального размера кадра. Использование кадров минимального размера подразумевает большой объём накладных расходов: каждый из кадров данных имеет служебный заголовок, размер которого не зависит от размера самого кадра.

Например, длина служебного заголовка у кадров длиной 64 байта (рис. 4б) и 156 байт (рис. 4в) будет одинакова, но объём пользовательских данных будет отличаться. Для того, чтобы передать 138 байт данных пользователя потребуется три кадра длиной 64 байт или один кадр длиной 156 байт, таким образом в первом случае потребуется передать 192 байта, а во втором - 156 байт. При одинаковой пропускной способности канала связи, использование кадров большого размера повысит эффективность, увеличив полезную пропускную способность системы. Данные о значениях производительности устройств Инфинет для различных условий представлены в документе Производительность устройств Инфинет.

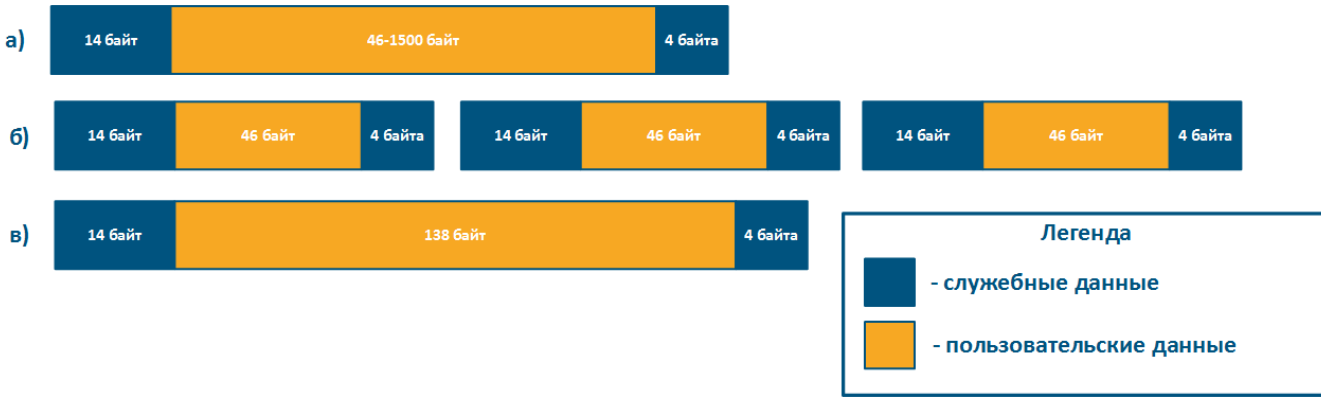


Рисунок 4 - Примеры структуры кадров Ethernet различной длины

Задержка

Под задержкой понимается время передачи пакета данных от источника до получателя. Величина задержки складывается из следующих компонентов:

- **Время распространения сигнала в среде:** зависит от физических характеристик среды, но, в любом случае, является ненулевым.
- **Время сериализации:** преобразование входящими/исходящими интерфейсами битового потока в сигнал и обратно не является мгновенным и требует аппаратных ресурсов сетевого устройства.
- **Время обработки:** время, которое пакет данных находится в устройстве. Это время зависит от состояния очередей сообщений, т.к. пакет данных будет обработан только после обработки пакетов, помещённых в эту очередь ранее.

При измерениях задержки часто используется понятие круговой задержки (RTT), т.е. времени распространения пакета данных от источника к получателю и обратно. Такое значение, например, используется при выводе результатов команды ping. Состояние промежуточных сетевых устройств при обработке прямого и обратного пакета данных может отличаться, поэтому в общем случае круговая задержка не равна двум односторонним задержкам.

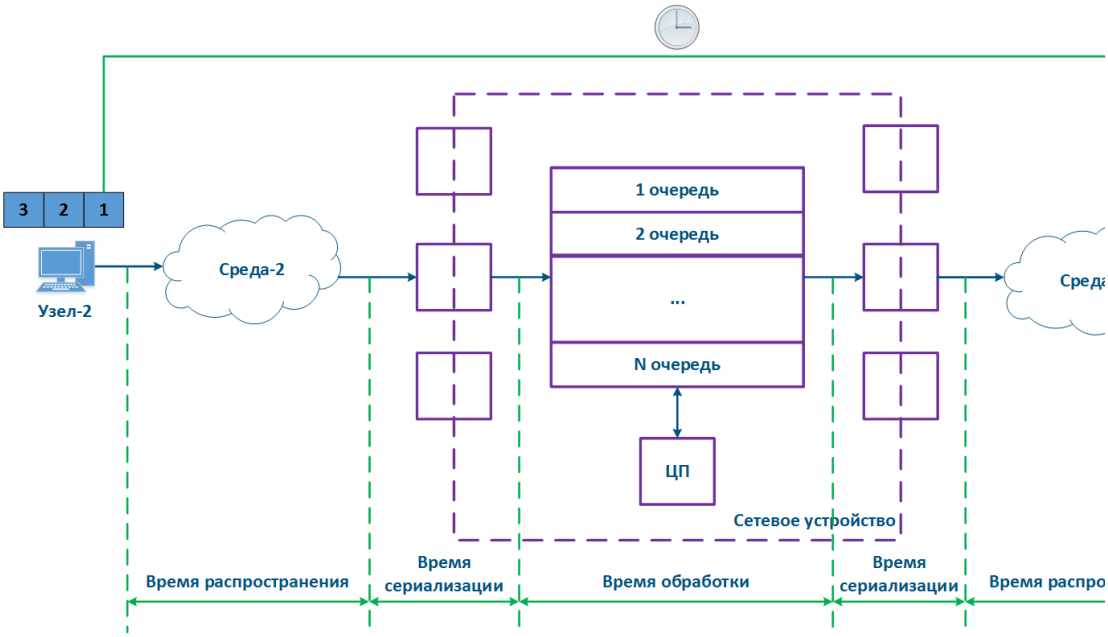


Рисунок 5 - Пример задержки при передаче данных

Джиттер

Загрузка ЦП и состояние очередей сообщений на промежуточных сетевых устройствах постоянно меняются, поэтому задержка при распространении пакетов данных может изменяться. В примере (рис. 6) время распространения пакетов с идентификаторами 1 и 2 отличаются. Разница между максимальным и средним значениями задержки называется джиттером.

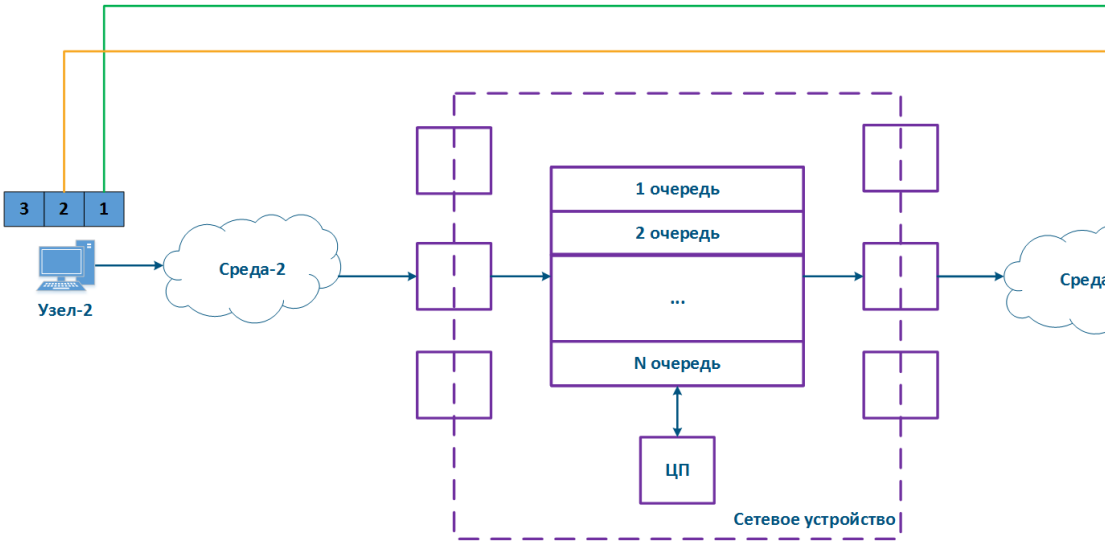


Рисунок 6 - Пример плавающей задержки при передаче данных

В сетевой инфраструктуре с избыточностью каналов связи данные между источником и получателем могут быть переданы различными путями, что, также, приведёт к появлению джиттера. В частном случае, разница между задержками в каналах связи может оказаться настолько большой, что порядок переданных пакетов данных изменится на приёмной стороне (рис. 7). В примере пакеты с идентификаторами были приняты в разном порядке.

Влияние эффекта зависит от характеристик сервиса и возможностей восстановления исходной последовательности протоколами высших уровней сетевого взаимодействия. Например, если трафик различных сервисов будет передан разными путями, то это не повлияет на неупорядоченность принятых данных.

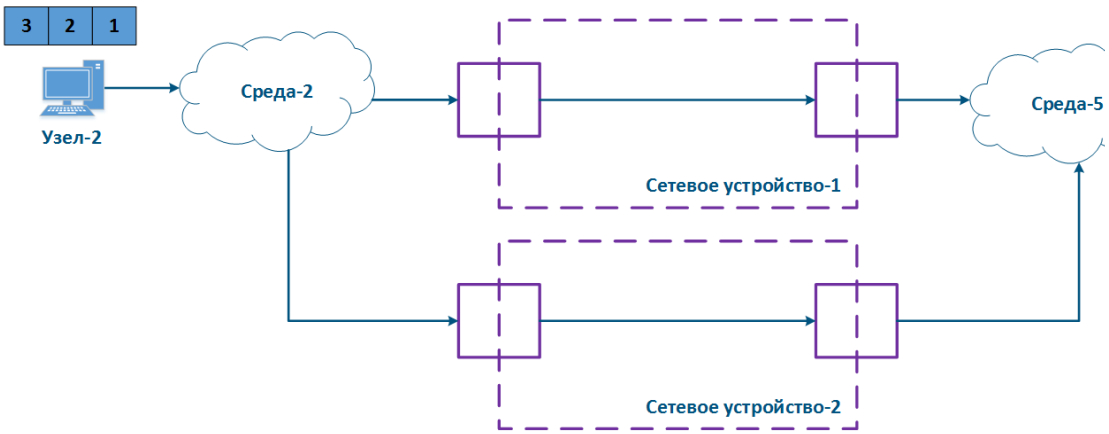


Рисунок 7 - Пример неупорядоченной доставки данных

Требования сервисов к метрикам качества

Каждый из сервисов передачи данных имеет набор требований к показателям качества. Документ RFC 4594 предусматривает следующие виды сервисов:

Сервис	Величина		
	Потери	Задержка	Джиттер
Трафик служебных протоколов	низкие	низкая	низкий
Телефония	очень низкие	очень низкая	очень низкий
Сигнализация	низкие	низкая	низкий
Мультимедийные конференции	средние	очень низкая	низкий
Интерактивный трафик реального времени	низкие	очень низкая	низкий
Мультимедийные трансляции	средние	средняя	низкий

Широковещательное видео	очень низкие	средняя	низкий
Данные, требовательные к задержкам	низкие	средняя	очень низкий
Управление	низкие	средняя	средний
Данные, требовательные к пропускной способности	низкие	высокая	высокий
Стандарт	не определено		
Низкоприоритетные данные	высокие	высокая	высокий

Application Categories	Service Class	Signaled	Flow Behavior	G.1010 Rating
Application Control	Signaling	Not applicable	Inelastic	Responsive
Media-Oriented	Telephony	Yes	Inelastic	Interactive
	Real-Time Interactive	Yes	Inelastic	Interactive
	Multimedia Conferencing	Yes	Rate Adaptive	Interactive
	Broadcast Video	Yes	Inelastic	Responsive
	Multimedia Streaming	Yes	Elastic	Timely
Data	Low-Latency Data	No	Elastic	Responsive
	High-Throughput Data	No	Elastic	Timely
	Low-Priority Data	No	Elastic	Non-critical
Best Effort	Standard	Not Specified		Non-critical

Методы обеспечения QoS

Передача трафика различных сервисов реализована на единой сетевой инфраструктуре, которая имеет ограниченные ресурсы, поэтому должны быть предусмотрены механизмы по распределению ресурсов между сервисами.

Рассмотрим пример (рис. 8), в котором Узел-2 генерирует трафик нескольких сервисов с суммарной скоростью 1 Гбит/с, Среда-2 позволяет передать этот поток данных промежуточному сетевому устройству, однако максимальная пропускная способность канала связи сетевого устройства и Узла-5 равна 500 Мбит/с. Очевидно, что поток данных не может быть обработан полностью и часть этого потока должна быть отфильтрована. Задача QoS сделать эту фильтрацию управляемой, обеспечив конечным сервисам требуемые значения метрик. Разумеется, не получится обеспечить требуемые показатели для всех сервисов, т.к. пропускные способности каналов связи не совпадают, поэтому в рамках реализации политики QoS трафик критичных сервисов должен обрабатываться в первую очередь.

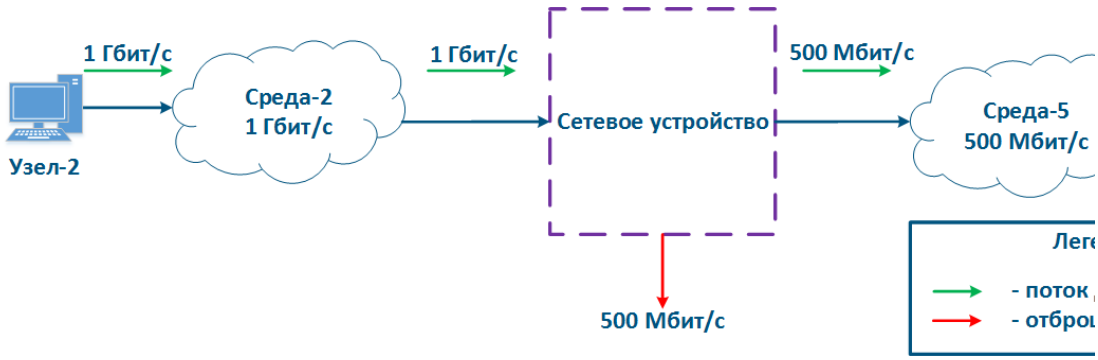


Рисунок 8 - Пример несогласованности объёма входящего трафика и пропускной способности каналов св

Рассмотренный пример позволяет сформулировать два основных метода, используемых при реализации политики QoS:

- **Приоритизация:** распределение данных по очередям сообщений и приоритетная выборка пакетов из очередей. В этом случае сначала обрабатываются пакеты наиболее чувствительные к задержке и джиттеру, а потом - трафик, для которого значение задержки не критично.
- **Ограничение пропускной способности:** ограничение пропускной способности для потоков трафика. Весь трафик, превышающий установленный порог пропускной способности, будет отброшен.

Рассмотрим пример, представленный выше, добавив в схему распространения данных второе промежуточное сетевое устройство (рис. 9а). Схема распространения пакетов описывается следующими этапами:

- **Этап 1:**

- Узел-1 и Узел-2 формируют пакеты двух сервисов: телефонии и почты. Трафик телефонии, в отличие от данных почтового сервиса, чувствителен к задержке и джиттеру (см. Требования сервисов к метрикам качества), поэтому должен быть обработан промежуточными устройствами в первую очередь.
- Сетевое устройство-1 принимает пакеты Узла-1 и Узла-2.

- **Этап 2:**
  - На Сетевом устройстве-1 настроена приоритизация трафика, которая заключается в том, что устройство классифицирует входящий трафик и помещает пакеты данных в различные очереди сообщений. Весь трафик телефонии будет помещён в очередь 0,а трафик почтового сервиса - в очередь 16. Таким образом, приоритет очереди 0 выше, чем очереди 16.
  - Освобождение очередей сообщений и передача данных исходящим интерфейсам осуществляются в соответствии с приоритетами очередей, т.е. сначала будет опустошена очередь 0, а затем - очередь 16.
- **Этап 3:**
  - Сетевое устройство-1 отправляет данные в Среду-7, связанную с Сетевым устройством-2. Последовательность пакетов данных соответствует метрикам качества - в первую очередь в среду переданы данные телефонии, а затем - почтового сервиса.
  - Узел-3 подключен к Сетевому устройству-2 и формирует поток данных почтового сервиса.
- **Этап 4:**
  - На Сетевом устройстве-2 отсутствуют настройки приоритизации, поэтому весь входящий трафик будет помещён в очередь сообщений 16. Отправка данных из очередей будет соответствовать последовательности их приёма, т.е. трафик телефонии будет обработан наравне в трафиком почтового сервиса, несмотря на требования к значениям метрик качества.
  - Сетевое устройство-2 вносит задержку во время распространения трафика телефонии.
- **Этап 5:**
  - Данные отправляются конечным узлам. Время распространения пакетов телефонии было увеличено за счёт обработки трафика почтового сервиса Узла-3.

Каждое из промежуточных сетевых устройств, на котором отсутствуют настройки приоритизации трафика, будет задерживать распространение данных, при этом величина вносимой задержки будет непредсказуемой. Таким образом, большое число промежуточных устройств сделает невозможным работу сервисов реального времени из-за недостижимости качественных метрик, т.е. настройка приоритизации трафика должна быть выполнена на всём пути распространения трафика в сети (рис. 9б).

Следует иметь в виду, что реализация политик QoS является одним из компонентов по обеспечению метрик качества. Для получения максимального эффекта конфигурация QoS должна быть синхронизирована с другими настройками. Например, использование технологии TDMA вместо Polling на устройствах семейств InfiLINK 2x2 и InfiMAN 2x2 позволяет снизить джиттер, стабилизировав значение задержки (см. TDMA и Polling: особенности применения в беспроводных сетях).

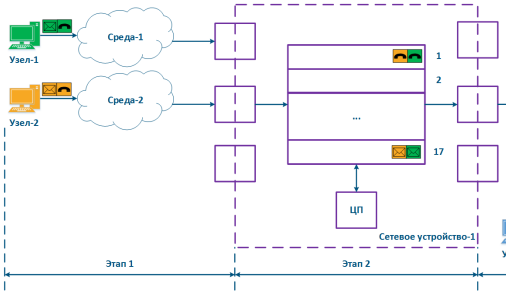


Рисунок 9а - Пример ра

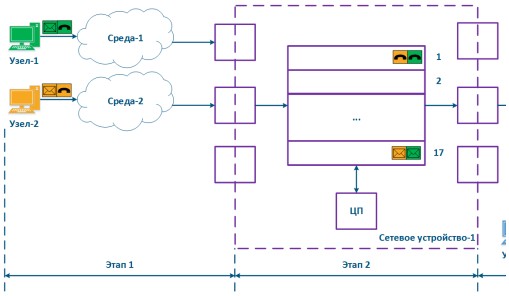


Рисунок 9б - Прим

Механизмы приоритизации трафика

С точки зрения возможности управления путь распространения трафика в сети может быть описан двумя концепциями (рис. 10а,б):

- **Белый ящик:** все сетевые устройства на пути данных находятся в одной зоне ответственности. В этом случае, конфигурация QoS на устройствах может быть согласована, что соответствует требованию, описанному в разделе выше.

- **Черный ящик:** часть сетевых устройств на пути данных находятся в чужой зоне ответственности. Правила классификации входящих данных и алгоритм выборки сообщений из очередей настраивается на каждом устройстве индивидуально. Это обусловлено тем, что реализация архитектуры очередей сообщений зависит от производителя оборудования, поэтому отсутствует гарантия корректной конфигурации QoS на устройствах в чужой зоне ответственности, и как следствие, отсутствует гарантия выполнения качественных метрик.

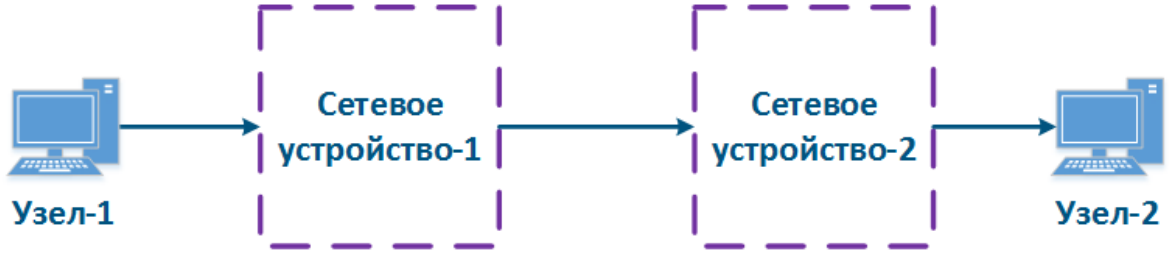


Рисунок 10а - Пример структуры "белого ящика"

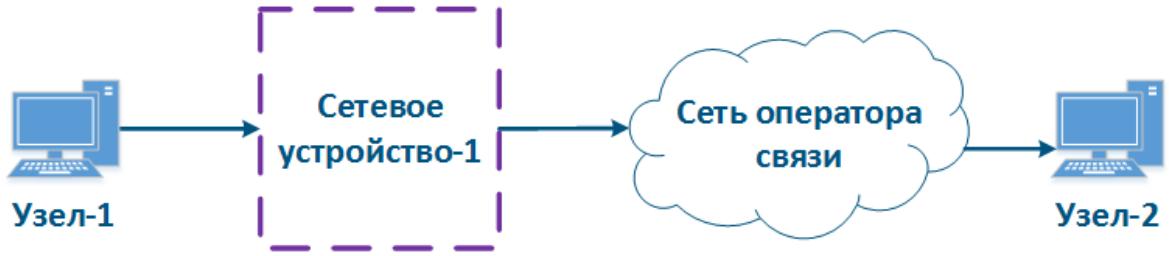


Рисунок 10б - Пример структуры "черного ящика"

Одним из решений описанной проблемы для сетевой структуры "черный ящик" является маркировка заголовков пакетов: приоритет, требуемый для обработки пакета, устанавливается в одном из полей заголовка и сохраняется на протяжении всего пути. В этом случае все промежуточные устройства могут помещать входящие данные в очередь сообщений в соответствии со значениями полей, где указан приоритет. Это потребует разработки стандартных протоколов и их реализации производителями оборудования.

Следует отметить, что в общем случае оборудование, находящееся в чужой зоне ответственности, не поддерживает приоритизацию данных в соответствии со значениями приоритета в служебных заголовков. Согласование приоритизации трафика на стыке зон ответственности должно быть выполнено на административном уровне.

Для установки приоритета обслуживания пакета могут использоваться служебные поля различных сетевых протоколов. В рамках данной статьи подробно рассмотрим использование заголовков протоколов Ethernet и IPv4.

Приоритизация в Ethernet (802.1p)

Заголовок кадров Ethernet включает в себя служебное поле "User Priority", которое предназначено для приоритизации кадров данных. Поле имеет размер 3 бита, что позволяет выделить 8 классов трафика: 0 класс - наименьший приоритет, 7 класс - наибольший приоритет. Следует иметь в виду, что поле "User Priority" присутствует только в кадрах 802.1q, т.е. тэгированных одной из меток VLAN.

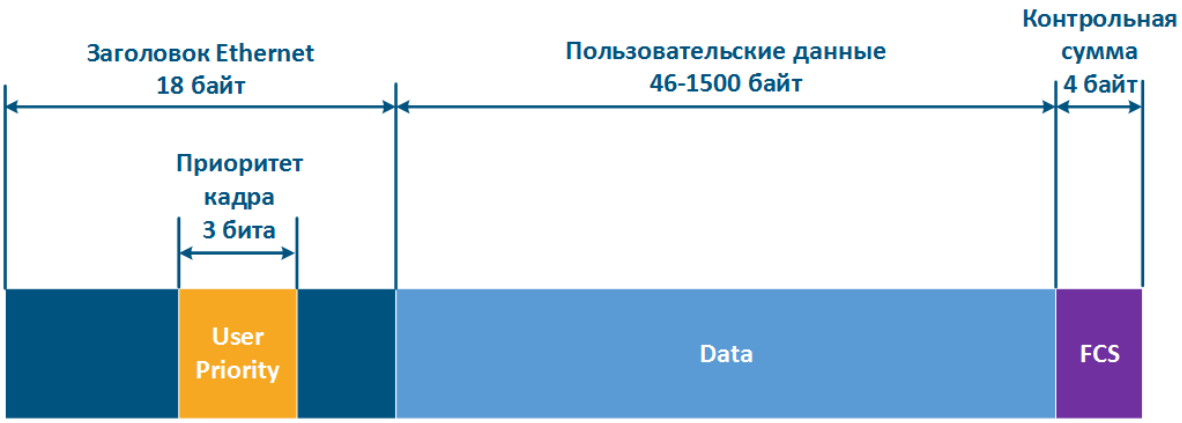


Рисунок 11 - Служебное поле в заголовке Ethernet для приоритизации кадров

Приоритизация в IP

Протокол IP включает в себя три исторических стадии развития служебного поля, отвечающего за приоритизацию пакетов:

1. В заголовке IP-пакета при утверждении протокола присутствовало поле ToS (Type of Service - тип сервиса) размером 8 бит (см. RFC 791). ToS включал в себя следующие поля (рис. 12а):
    a. Precedence: значение приоритета.
    b. Delay: бит минимизации задержки.
    c. Thorughput: бит максимизации пропускной способности.
    d. Reliability: бит максимизации надёжности.
    e. 2 бита, значения которых равны 0.
2. Для приоритизации пакетов по-прежнему использовались 8 битов, однако ToS теперь включал в себя следующие поля (см. RFC 1349):
    a. Delay.
    b. Throughput.
    c. Reliability.
    d. Cost: бит минимизации метрики стоимости (используется 1 бит, значение которого ранее было нулевым).
3. Структура заголовка IP была изменена (см. RFC 2474). 8 бит, используемые ранее для приоритизации, были распределены следующим образом (рис. 12б):
    a. DSCP (Differentiated Services Code Point - код дифференцированной услуги): приоритет пакета.
    b. 2 бита зарезервировано.

Таким образом, ToS позволяет выделить 8 классов трафика: 0 - наименьший приоритет, 7 - наивысший приоритет, а DSCP - 64 класса: 0 - наименьший приоритет, 63 - наивысший приоритет.

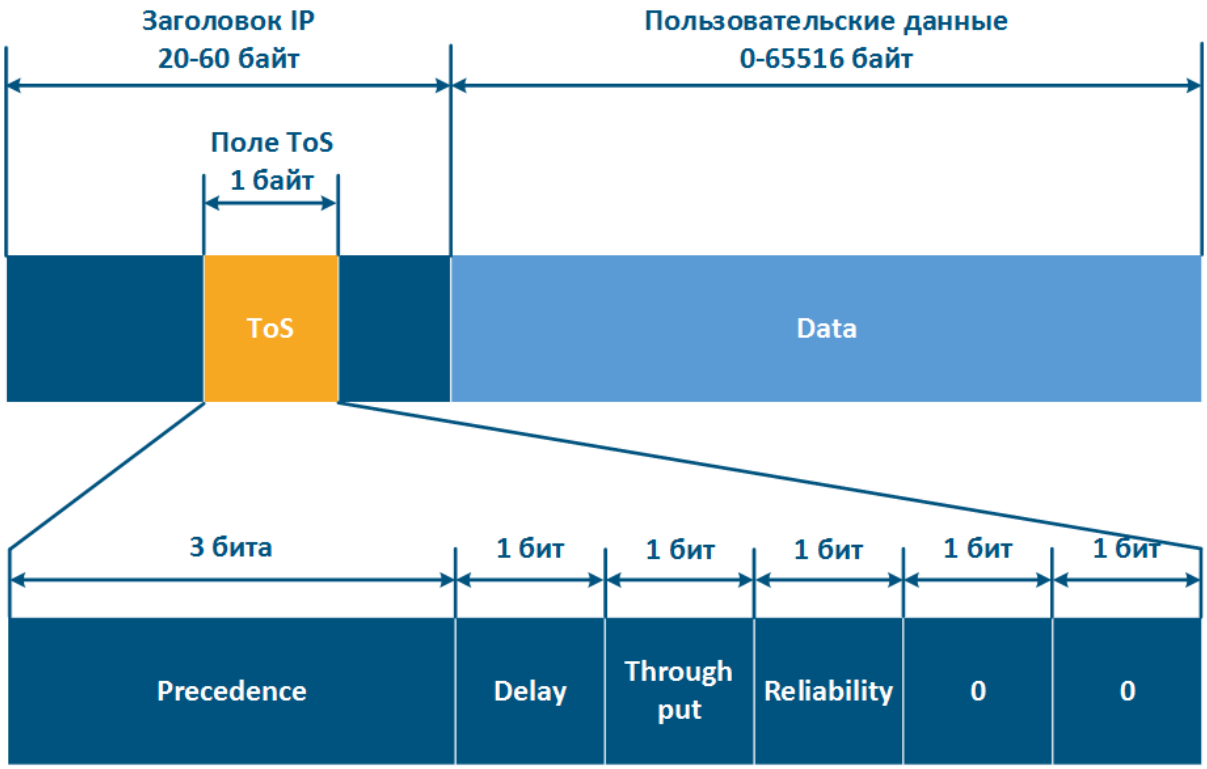


Рисунок 12а - Служебное поле ToS в заголовке IP для приоритизации пакетов

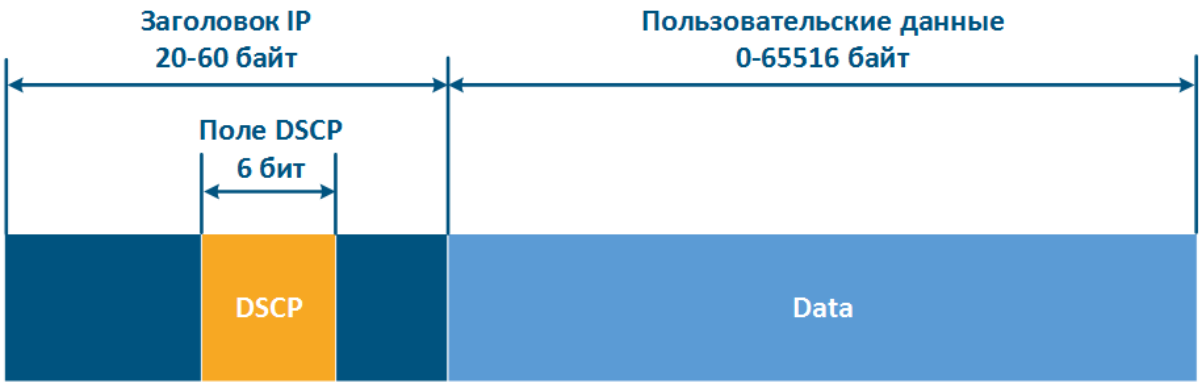


Рисунок 12б - Служебное поле DSCP в заголовке IP для приоритизации пакетов

Установка приоритета

Множество конечных узлов в сети не поддерживают операции по установке и удалению приоритетов в служебных заголовках, поэтому эта функциональность должна быть реализована в промежуточных сетевых устройствах

Рассмотрим пример распространения данных от Узла-1 к Узлу-2 через DS-домен и стороннюю сеть оператора связи (рис. 13а-в). DS-домен включает в себя три устройства, два из которых являются для домена пограничными, а одно - промежуточным. Рассмотрим этапы обработки данных в сети на примере передачи кадра Ethernet (основные принципы, рассмотренные в примере, применимы для IP-пакета или другого протокола, поддерживающего приоритизацию данных):

- **Этап 1:** Узел-1 формирует кадр Ethernet для Узла-2. Поле для установки приоритета кадра в заголовке отсутствует (рис. 13а).
- **Этап 2:** Пограничное сетевое устройство-1 меняет заголовок Ethernet, устанавливая в поле приоритета значение 1. На пограничных устройствах должны быть настроены правила для выборки трафика Узла-1 из общего потока, для того, чтобы необходимый приоритет был установлен только этим кадрам. В сетях с большим числом потоков трафика список правил на пограничных устройствах может быть объёмным. Пограничное сетевое устройство-1 обрабатывает кадр в соответствии с установленным приоритетом, помещая его в соответствующую очередь сообщений. Кадр передаётся на исходящий интерфейс и отправляется в сторону Промежуточного сетевого устройства-2 (рис. 13а).
- **Этап 3:** Промежуточное сетевое устройство-2 принимает кадр Ethernet, в котором установлен приоритет 1, и помещает его в соответствующую очередь сообщений. Устройство не выполняет манипуляций по установке/удалению приоритета в заголовке кадра. Кадр передаётся в сторону Пограничного сетевого устройства-3 (рис. 13а).
- **Этап 4:** Пограничное сетевое устройство-3 обрабатывает входящий кадр аналогично Промежуточному устройству-2 (см. Этап 3) и передаёт его в сторону Сети оператора связи(рис. 13а).
    - **Этап 4б:** в случае, если существует договорённость о том, что трафик будет передан через Сеть оператора связи с приоритетом, отличным от 1, то Пограничное устройство-3 должно выполнить изменение приоритета. В рассматриваемом примере, устройства меняет значение приоритета с 1 на 6 (рис. 13б).
- **Этап 5:** при распространении кадра через Сеть оператора связи устройства руководствуются значением приоритета в заголовке Ethernet (рис. 13а).
    - **Этап 5б:** аналогично Этапу 5 (рис. 13б).
    - **Этап 5в:** при отсутствии договорённости о приоритизации кадров данных в соответствии со значением приоритета, указанным в заголовке Ethernet, сторонний оператор связи может применить к трафику собственную политику QoS и установить приоритет, который может не согласовываться с политикой QoS, принятой в DS-домене (рис. 13в).
- **Этап 6:** пограничное устройство в Сети оператора связи удаляет поле приоритета из заголовка Ethernet и передаёт его в направлении Узла-2 (рис. 13а-в).

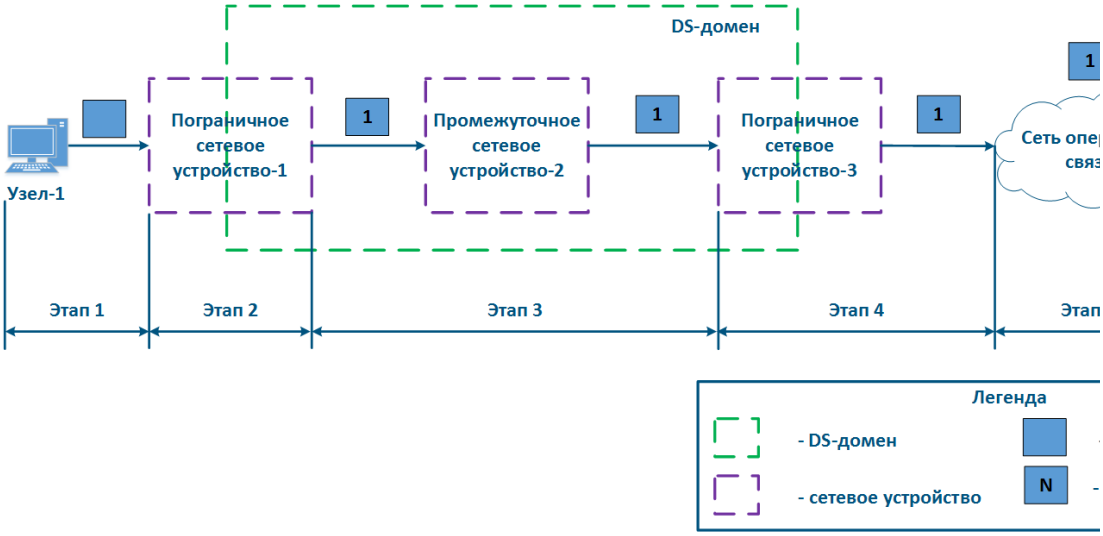


Рисунок 13а - Пример изменения приоритета кадра Ethernet при распространении через два сегмента сети (приоритет се

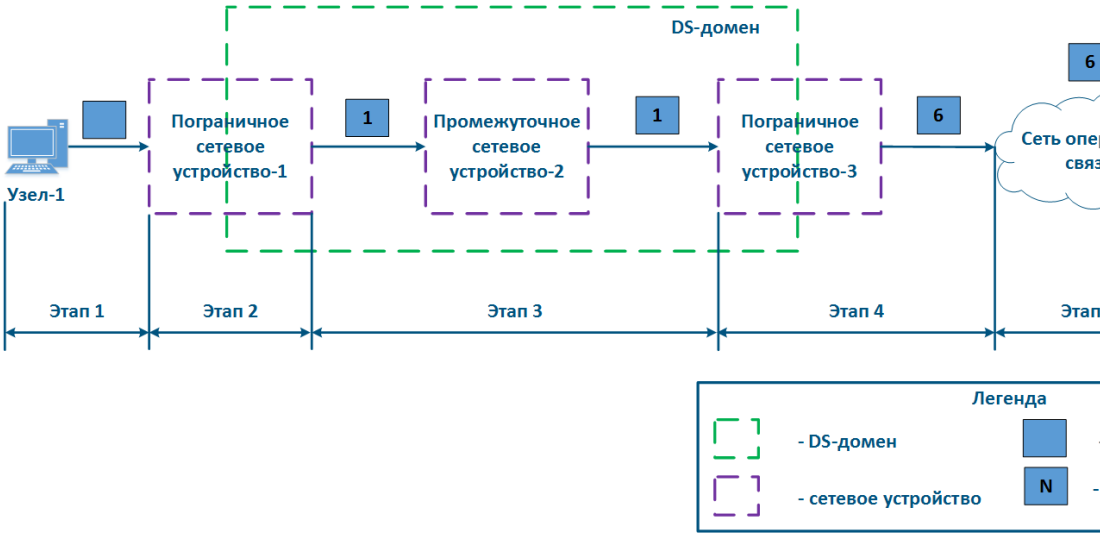


Рисунок 13б - Пример изменения приоритета кадра Ethernet при распространении через два сегмента сети (приоритет сегментов согл

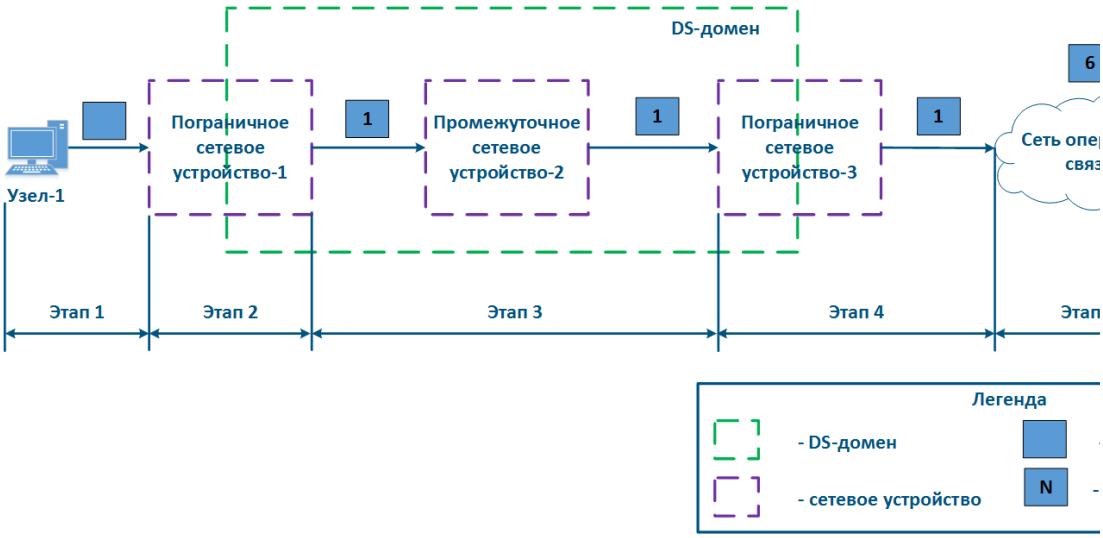


Рисунок 13в - Пример изменения приоритета кадра Ethernet при распространении через два сегмента сети (приоритет сег

Реализация очередей в устройствах Инфинет

Процесс анализа устройством приоритета в служебных заголовках и обработка данных в соответствии с этими приоритетами не является простым по следующим причинам:

- Устройства поддерживают автоматическое распознавание приоритетов в соответствии с разными протоколами. Например, устройства семейства InfiLINK XG поддерживают распознавание приоритетов 802.1p и не распознаёт значения приоритетов DSCP.
- Устройства, являющиеся пограничными для DS-домена, позволяют использовать разный набор критериев для классификации трафика. Например, устройства InfiMAN 2x2 позволяют установить приоритет, выбрав весь TCP-трафик, направленный на порт 23, устройства семейства Vector 5 - нет.
- Число очередей, реализованных в устройствах разных производителей, отличается. Для того, чтобы установить соответствие между приоритетом в служебном заголовке данных и внутренней очередью устройства, должна быть использована таблица соответствия.

Данные о внутренней архитектуре очередей, возможностях управления приоритетами данных и соответствия между протокольными и внутренними значениями приоритетов приведены в таблицах ниже.

Следует отметить архитектурную особенность организации очередей в устройствах Инфинет: все очереди делят между собой единый буфер памяти. В случае, если весь трафик попадает в одну очередь, то её размер будет соответствовать размеру буфера, а если очередей будет несколько, то размер буфера памяти будет равномерно поделен между ними.

Таблица внутренней организации очередей сообщений

Параметр	Описание	InfiLINK 2x2 / InfiMAN 2x2	InfiLINK Evolution / InfiMAN Evolution	InfiLINK XG / InfiLINK XG 1000	Vector 5 / Vector 6 / Vector 70
Критерий маркировки	Набор критериев, которые могут использоваться при классификации входящего трафика.	поддержка PCAP-выражений (PCAP выражения позволяют выполнить гибкую фильтрацию на основе любых полей служебных заголовков, см. PCAP-фильтры)	поддержка PCAP-выражений (PCAP выражения позволяют выполнить гибкую фильтрацию на основе любых полей служебных заголовков, см. PCAP-фильтры)	vlan-id	vlan-id
Автораспознавание	Для указанных протоколов семейство устройств позволяет выполнить автоматическое распознавание приоритета, установленного в заголовке и помещение данных в соответствующую очередь.	RTP 802.1p IPIP/GRE-туннели MPLS DSCP ToS ICMP TCP Ack PPPoE	RTP 802.1p IPIP/GRE-туннели MPLS DSCP ToS ICMP TCP Ack PPPoE	802.1p	802.1p
Число очередей	Количество очередей сообщений, используемое в устройстве.	17	17	4	8
Диспетчеризация очередей	Поддерживаемые механизмы выборки сообщений из очередей сообщений.	Взвешенная	Взвешенная	Строгая Взвешенная	
Настройка приоритизации в Web	Ссылки на документацию по настройке приоритизации трафика через Web-интерфейс.	Параметры QoS Контроль трафика	Параметры QoS Контроль трафика	Настройка QoS Раздел Коммутатор Коммутация на	Настройка коммутации

				основе VLAN	
Настройка приоритизации в CLI	Ссылки на документацию по настройке приоритизации трафика через интерфейс командной строки.	Команда qm	Команда qm	Команды настройки коммутатора	-

Таблица соответствия протокольных и внутренних приоритетов для устройств семейств InfiLINK 2x2, InfiMAN 2x2, InfiLINK Evolution, InfiMAN Evolution

Класс трафика (в соответствии с MINT)	InfiLINK 2x2, InfiMAN 2x2, InfiLINK Evolution, InfiMAN Evolution	802.1p	ToS (Precedence)	DSCP
Background	16	01		
Regular best effort	15	00	00	0
Business 6	14		01	8, 10
Business 5	13			12, 14
Business 4	12		02	16, 18
Business 3	11			20, 22
Business 2	10		03	24, 26
Business 1	9	02		28, 30
QoS 4	8		04	32
QoS 3	7			34
QoS 2	6			36
QoS 1	5	03		38
Video 2	4	04	05	40, 42
Video 1	3			44, 46
Voice	2	05	06	48, 50
Control	1	06		52, 54
NetCrit	0	07	07	56, 58, 60, 62

Таблица соответствия протокольных и внутренних приоритетов для устройств семейств InfiLINK XG, InfiLINK XG 1000, Vector 5, Vector 6, Vector 70

Класс трафика (в соответствии с 802.1p)	802.1p	InfiLINK XG, InfiLINK XG 1000	Vector 5, Vector 6, Vector 70
Background (наименьший приоритет)	00	1	0
Best Effort	01		1
Excellent Effort	02	2	2
Critical Applications	03		3
Video	04	3	4
Voice	05		5
Internetwork Control	06	4	6
Network Control (наивысший приоритет)	07		7

Диспетчеризация очередей

Приоритизация сообщений подразумевает под собой использование нескольких очередей сообщений, содержимое которых должны быть передано исходящим интерфейсам через единую шину сообщений. Устройства Инфинет поддерживают два механизма передачи сообщений из очередей в шину: строгая и взвешенная диспетчеризация.

Строгая диспетчеризация

Механизм строгой приоритизации подразумевает последовательное опустошение очередей в соответствии со значениями приоритета. Отправка сообщений с приоритетом 2 будет выполнена только после того, как в шину будут переданы все сообщения с приоритетом 1 (рис. 14). После того, как будут отправлены сообщения с приоритетами 1 и 2, устройство начнёт отправку сообщений с приоритетом 3.

Данный механизм имеет явный недостаток: низкоприоритетному трафику не будут выделяться ресурсы, если есть сообщения в более приоритетных очередях, что приведёт к полной недоступности некоторых сетевых сервисов.

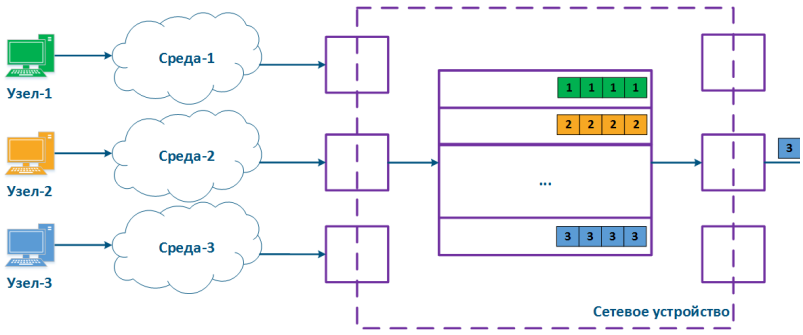


Рисунок 14 - Строгая диспетчеризация сс

Взвешенная диспетчеризация

Взвешенная диспетчеризация лишена недостатков строгой диспетчеризации. Взвешенная диспетчеризация подразумевает распределение ресурсов между всеми очередями сообщений в соответствии с весовыми коэффициентами, которые соответствуют значениям приоритета. В случае трёх очередей сообщений (рис. 15), весовые коэффициенты могут быть распределены следующим образом:

- очередь сообщений 1: вес = 3;
- очередь сообщений 2: вес = 2;
- очередь сообщений 3: вес = 1.

При использовании взвешенной диспетчеризации каждая из очередей сообщений получит ресурсы, т.е. не возникнет ситуации с полной недоступностью одного из сетевых сервисов.

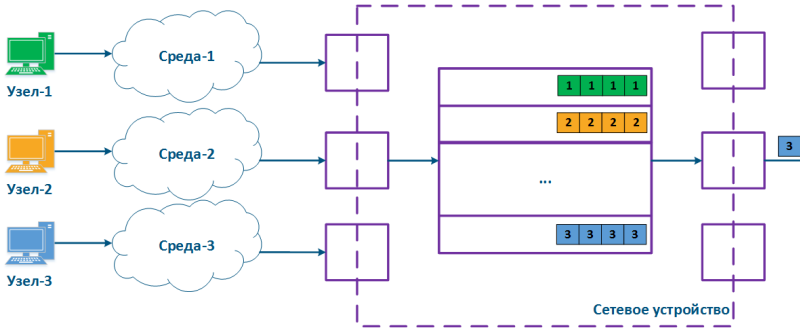


Рисунок 15 - Взвешенная диспетчеризация

Рекомендации по приоритизации трафика

Можно сформулировать набор универсальных рекомендаций по конфигурации механизмов приоритизации трафика:

- Необходимо скрупулезно отнестись к разработке политики QoS. Политика должна описывать трафик всех сервисов, используемых в сети, предусматривать строгое соответствие сервиса и класса трафика.
- Политика QoS должна учитывать технические возможности устройств по распознаванию и манипуляции со значениями служебных полей, в которых указывается приоритет данных.
- На пограничных устройствах DS-домена должны быть настроены правила классификации потоков трафика.
- На промежуточных устройствах DS-домена должна быть активирована функция автоматического распознавания приоритетов трафика.

Механизмы ограничения пропускной способности

Распределение ресурсов сети между потоками трафика может быть выполнено не только за счёт приоритизации, но и с помощью механизма ограничения пропускной способности. В этом случае, скорость передачи данных потока не может превысить пороговый уровень, установленный администратором сети.

Принцип ограничения скорости в устройствах Инфинет

Принцип ограничения скорости заключается в постоянном измерении интенсивности потока данных и, в случае, если значение интенсивности превышает установленный порог, срабатывает ограничение (рис. 16а,б). Для ограничения пропускной способности в устройствах Инфинет используется алгоритм Token Bucket, заключающийся в том, что все пакеты данных сверх порога пропускной способности отбрасываются. В результате образуются потери, описанные выше.

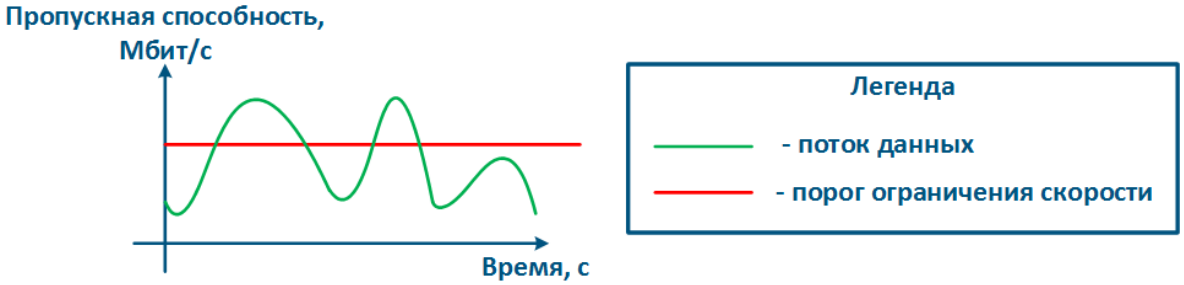


Рисунок 16а - График интенсивности потока данных без ограничения

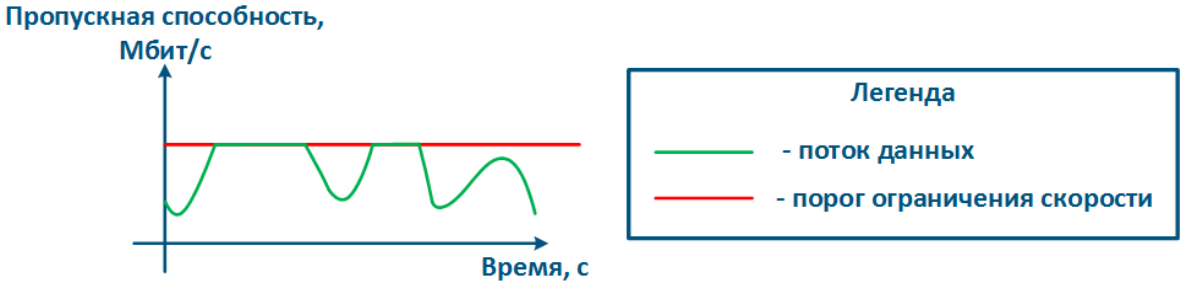


Рисунок 16б - График интенсивности потока данных после ограничения

Алгоритм Token Bucket

Для каждого правила ограничения скорости формируется логический буфер, содержащий объём разрешённых для передачи данных. Как правило, размер этого буфера больше, чем размер ограничений. Каждую единицу времени такому буферу выделяется размер данных, равный установленному порогу ограничения скорости.

В рассматриваемом примере (видеоролик 1) ограничение скорости составляет 3 единицы данных, размер буфера - 12 единиц данных. Буфер постоянно пополняется в соответствии с установленным порогом, однако не может быть заполнен больше собственного объёма.

Your browser does not support the HTML5 video element

Видеоролик 1 - Выделение ресурсов буферу ограничения скорости

Обработка данных, поступивших на входящий интерфейс устройства, будет выполнена только в том случае, если буфер содержит ресурсы для их обработки (видеоролик 2). Таким образом прохождение данных опустошает буфер ресурсов. Если в момент прихода данных буфер будет пуст, то данные будут отброшены.

Your browser does not support the HTML5 video element

Видеоролик 2 - Использование выделенных ресурсов при обработке данных

Следует понимать, что процессы выделения ресурсов буферу ограничения скорости и обработки данных выполняются одновременно (видеоролик 3).

Интенсивность потоков данных в пакетных сетях непостоянна, что позволяет проявить одно из достоинств алгоритма Token Bucket. Интервалы времени, в которые не передаются данные, позволяют выполнить накопление ресурсов в буфере, а затем обработать объём данных, превышающий порог ограничения. Импульсным потокам данных, например web-трафик, будет выделена широкая полоса, позволяющая выполнить быструю загрузку web-страниц, повысив уровень комфорта конечного пользователя.

Несмотря на описанное преимущество алгоритма Token Bucket, средняя пропускная способность будет соответствовать установленному порогу, т.к. на длительных интервалах времени объём ресурсов будет определяться не размером буфера, а интенсивностью его заполнения, которая соответствует порогу пропускной способности.

Your browser does not support the HTML5 video element

Видеоролик 3 - Обработка данных буфером ограничения скорости

Алгоритм Token Bucket может быть применён для отдельных потоков трафика, в этом случае буфер ограничения скорости будет выделен для каждого из потоков (видеоролик 4).

В рассматриваемом примере создано два правила ограничения скорости: для трафика vlan 161 - 3 единицы данных в единицу времени, для трафика vlan 162 - 2 единицы данных. Размер буфера для каждого из потоков трафика равен 4 интервалам времени, т.е. 12 единиц данных для трафика vlan 161 и 8 единиц данных для трафика vlan 162. Суммарно буферам выделяется 5 единиц данных в каждый из интервалов времени, далее выделенные ресурсы распределяются между буферами. Поскольку размер буферов ограничен, то ресурсы, выделенные сверх их размеров, не могут быть использованы.

Your browser does not support the HTML5 video element

Видеоролик 4 - Выделение ресурсов двум буферам ограничения скорости

Ресурсы каждого буфера могут быть использованы только для трафика соответствующего сервиса (видеоролик 5). Так, для обработки трафика vlan 161 используется буфер ресурсов для трафика vlan 161. Аналогично используются ресурсы буфера для трафика vlan 162.

Your browser does not support the HTML5 video element

Видеоролик 5 - Использование выделенных ресурсов для обработки данных

Существуют способы связи буферов ресурсов друг с другом. Например, в устройствах Инфинет буферы выделенных ресурсов могут быть связаны через классы (см. ниже). В случае, если один из буферов ресурсов будет заполнен (видеоролик 6), выделенные ему ресурсы могут быть предоставлены другому буферу.

В примере буфер для трафика vlan 162 заполнен, что позволяет пополнить буфер трафика vlan 161 выделенными 5 единицами данных, вместо 3. В этом случае пропускная способность сервиса vlan 161 вырастет. Но как только буфер ресурсов трафика vlan 162 появится свободная память, то распределение ресурсов вернётся к нормальному режиму: трафику vlan 161 - 3 единицы данных, трафику vlan 162 - 2 единицы данных.

Your browser does not support the HTML5 video element

Видеоролик 6 - Перераспределение выделенных ресурсов между буферами ограничения трафика различных сервисов

Виды ограничений скорости в устройствах Инфинет

Рассмотренный принцип ограничения пропускной способности реализован в устройствах Инфинет двумя способами:

- Ограничение трафика физического интерфейса: ограничение будет применено к суммарному трафику всех потоков данных, проходящих через физический интерфейс. Данный метод прост в конфигурации - следует указать интерфейс и значение порога, но не позволяет применить ограничение к трафику конкретного сетевого сервиса.
- Ограничение потока трафика: ограничение применяется к логическому потоку данных. Логический поток данных выделяется из общего трафика с помощью проверки на соответствие заданным критериям, что позволяет применять ограничения пропускной способности к трафику сетевых сервисов, выбор которых выполняется на основе значений полей служебных заголовков. Например, можно выделить в логический канал весь трафик с vlan 42 и ограничить только его пропускную способность.

Устройства Инфинет позволяют настраивать иерархические структуры распределения ресурсов пропускной способности. Для этого используются два типа объектов: логический канал и класс, которые связаны отношением "потомок-родитель" соответственно. В классе указывается пропускная способность, которая будет распределена между дочерними логическими каналами, а в канале значения гарантированной и максимальной пропускной способностей, CIR и MIR соответственно.

Рассмотрим пример передачи трафика двух сервисов, ассоциированных с идентификаторами vlan 161 и 162, между Master и Slave (рис. 17а). Суммарному трафику сервисов выделено не более 9 Мбит/с.

Конфигурация устройства Master может быть выполнена следующим образом (рис. 17б):

- Создан Класс 16, пропускная способность которого ограничена значением 9 Мбит/с.
- Класс 16 является родительским по отношению к каналам 161 и 162, т.е. сумма трафика этих логических каналов ограничена значением 9 Мбит /с.
- Трафик с идентификатором vlan 161 ассоциируется с логическим каналом 161, vlan 162 - с логическим каналом 162.
- Значения CIR для канала 161 равно 4 Мбит/с, канала 162 - 5 Мбит/с. Если оба сервиса будут активно обмениваться данными, то пороговые значения для их трафика составят CIR, установленные для каждого из каналов.
- Значения MIR для канала 161 равно 9 Мбит/с, канала 162 - 7 Мбит/с. Если трафик в логическом канале 162 будет отсутствовать, то пороговое значение для канала 161 будет равно MIR, т.е. 9 Мбит/с. В обратном случае, пороговое значение для канала 162 будет равно 7 Мбит/с.

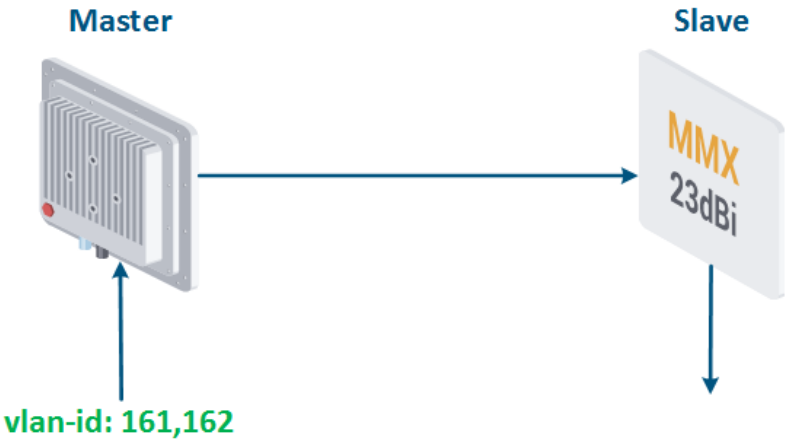
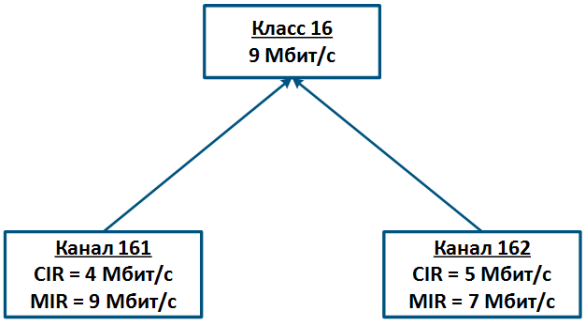


Рисунок 17а - Пример ограничения пропускной способности для трафика с vlan-id = 161, 162



Сценарий	Пропу
	vlan 16
• трафик vlan 161: присутствует • трафик vlan 162: присутствует	4 Мбит/
• трафик vlan 161: присутствует • трафик vlan 162: отсутствует	9 Мбит/
• трафик vlan 161: отсутствует • трафик vlan 162: присутствует	-

Рисунок 17б - Иерархическая структура каналов ограничения пропускной способности для трафика с vlan 16

Функциональные возможности по конфигурации ограничения пропускной способности на устройствах Инфинет всех семейств представлены в таблице ниже:

Таблица функциональных возможностей по ограничению пропускной способности в устройствах Инфинет

Параметр	Описание	InfiLINK 2x2 / InfiMAN 2x2	InfiLINK Evolution / InfiMAN Evolution	InfiLINK XG / InfiLINK XG 1000
Ограничение на интерфейсе	Возможность ограничения пропускной способности для физического интерфейса устройства.	-	-	• GE0 • GE1 • SFP • mgmt
Ограничение логического потока	Возможность ограничения пропускной способности для потока трафика, выделенного по одному или нескольким критериям.	до 200 логических каналов	до 200 логических каналов	-
Направление трафика	Возможность применения ограничений к входящему/исходящему потокам трафика.	входящий и исходящий	входящий и исходящий	исходящий
Иерархия ограничений	Возможность создания системы взаимных иерархических ограничений.	до 200 классов, являющихся дочерними по отношению к логическим каналам	до 200 классов, являющихся дочерними по отношению к логическим каналам	-
Критерии правил логических потоков	Критерии, используемые для выделения потоков данных.	поддержка PCAP-выражений (PCAP выражения позволяют выполнить гибкую фильтрацию на основе любых полей служебных заголовков, см. PCAP-фильтры)	поддержка PCAP-выражений (PCAP выражения позволяют выполнить гибкую фильтрацию на основе любых полей служебных заголовков, см. PCAP-фильтры)	-
Настройка	Ссылки на документацию по настройке	Контроль трафика	Контроль трафика	Раздел

ограничений в Web	ограничений пропускной способности через Web-интерфейс.			Коммутатор
Настройка ограничений в CLI	Ссылки на документацию по настройке ограничений пропускной способности через CLI.	Команда qm	Команда qm	Команды настройки коммутатора

Рекомендации по конфигурации ограничения пропускной способности

При конфигурации ограничения пропускной способности потоков данных следует руководствоваться следующими рекомендациями:

- Следует выполнять ограничение для трафика всех сетевых сервисов. Эти действия позволят сохранить контроль над всеми потоками трафика и осознанно выделять ресурсы для этих потоков.
- Ограничение пропускной способности должно выполняться на устройствах, расположенных ближе всего к источнику данных. Нет необходимости дублировать правила ограничения пропускной способности для потока данных на протяжении всей цепочки промежуточных устройств.
- Многие сетевые сервисы являются двунаправленными, что требует применения ограничений на устройствах как к входящему, так и исходящему трафику.
- Для корректной установки пороговых значений пропускной способности следует предварительно оценить средние и максимальные значения трафика сервисов. Особое внимание следует обратить на часы наибольшей нагрузки. Выполнить сбор данных для проведения анализа можно с использованием системы мониторинга InfiMONITOR.
- Сумма значений CIR логических каналов, ассоциированных с одним классом, не должна быть более максимальной пропускной способности класса.

Дополнительные материалы

White papers

1. TDMA и Polling: особенности применения в беспроводных сетях
2. Производительность устройств Инфинет

Вебинары

1. Настройка политик QoS в устройствах Инфинет.

Видео

1. Настройка политик QoS в устройствах Инфинет.

Прочее

1. RFC 4594.
2. RFC 791.
3. RFC 1349.
4. RFC 2474.
5. Система мониторинга InfiMONITOR.
6. Веб-интерфейс устройств семейств InfiLINK 2x2, InfiMAN 2x2. Параметры QoS.
7. Веб-интерфейс устройств семейств InfiLINK 2x2, InfiMAN 2x2. Контроль трафика.
8. Веб-интерфейс устройств семейств InfiLINK Evolution, InfiMAN Evolution. Параметры QoS.
9. Веб-интерфейс устройств семейств InfiLINK Evolution, InfiMAN Evolution. Контроль трафика.
10. Веб-интерфейс устройств семейств InfiLINK XG, InfiLINK XG 1000. Настройка QoS.
11. Веб-интерфейс устройств семейства Vector 5, Vector 6. Настройка коммутации.
12. Веб-интерфейс устройств семейства Vector 70. Настройка коммутации.
13. Настройка QoS manager в ОС WANFleX.