

Пример конфигурации RoCEv2 для коммутаторов серии QSW-6900





Оглавление

QOS SCHEDULING	3
1.1. Режим совместимости	3
1.2. Планирование перегрузки	3
1.3. Настройка значения COS для RDMA трафика	3
1.4. Режим доверия QOS, ограничение скорости очередей	4
2. MMU	4
2.1. MMU global	4
3. PFC	4
3.1. global headroom	4
3.2. Параметры MMU на Downlink интерфейсе и включение PFC	4
3.3. Параметры MMU на Uplink интерфейсе и включение PFC (если планирование RDMA трафика производится только на данном устройстве, следующая конфигурация не требуется)	5
3.4. Глобальное включение PFC	5
4. ECN	5
4.1. Настройка ECN на downlink интерфейсе	5
4.2. Настройка ECN на uplink интерфейсе	6
5. ОБЩАЯ ИНФОРМАЦИЯ	8
5.1. Замечания и предложения	8
5.2. Гарантия и сервис	8
5.3. Техническая поддержка	8



QOS SCHEDULING

1.1. Режим совместимости

```
qos-queue compatible enable
```

```
!
```

1.2. Планирование перегрузки

```
mls qos scheduler drr
```

```
!
```

```
drr-queue bandwidth 1 2 3 15 20 60 0 0
```

```
!
```

Рекомендуется настроить режим SP+DRR. Веса очереди корректируются по мере необходимости.

1.3. Настройка значения COS для RDMA трафика

```
expert access-list extended RoCEv2-Cnp
```

```
10 permit udp any any any any eq 4791 dscp 35
```

```
!
```

```
expert access-list extended RoCEv2-Data
```

```
10 permit udp any any any any eq 4791 dscp af41
```

```
!
```

```
class-map RoCEv2-Data
```

```
match access-group RoCEv2-Data
```

```
class-map RoCEv2-Cnp
```

```
match access-group RoCEv2-Cnp
```

```
policy-map RoCEv2
```

```
class RoCEv2-Data
```

```
set cos 5 priority
```

```
class RoCEv2-Cnp
```

```
set cos 6 priority
```

```
!
```

```
service-policy input RoCEv2
```

```
!
```

Соответствующий трафик ROCEv2 сопоставляется с очередями 5 и 6. Есть несколько способов реализовать это, а также можно напрямую использовать метод сопоставления dscp-cos. В примере конфигурации трафик RDMA проходит в очереди 5 и 6. Пакеты TCP идут в другие очереди. Фактической конфигурация может быть скорректирована. Приоритет конфигурации политики выше, чем у сопоставления dscp-cos.



1.4. Режим доверия QOS, ограничение скорости очередей

```
interface AggregatePort 1
 mls qos trust dscp
 qos queue 6 bandwidth maximum 12500000
!
```

Должен быть настроен режим порта trust dscp, на интерфейсах может быть включено ограничение скорости для очереди при необходимости. Для LAG конфигурация действует на порты-участники. Например, есть два порта 25G. Порты агрегации должны быть настроены на 12,5G для ограничения скорости до половины пропускной способности.

2. MMU

2.1. MMU global

```
mmu usage-warn-limit set 90
mmu sample-period buffer-counter 1
mmu sample-period queue-counter 1
!
```

Рекомендуемая конфигурация при необходимости. Сигнализация использования MMU, частота выборки буфера MMU и частота выборки очереди.

3. PFC

3.1. global headroom

```
mu pg-headroom set 9928
!
```

Рекомендуемые параметры конфигурации. Требуется отладки в реальных сценариях.

3.2. Параметры MMU на Downlink интерфейсе и включение PFC

MMU настраивается на физическом интерфейсе, а не на агрегированном интерфейсе.

```
interface TF GigabitEthernet 0/1
```

Настройки исходящей очереди трафика RDMA.

```
mmu queue-threshold output unicast 5 set 100
mmu queue-threshold output unicast 6 set 100
```

Настройка порога headroom для MMU.

```
mmu headroom-threshold pg 5 set 125
```

Настройка порога для входящего трафика.

```
mmu xoff-threshold pg 5 set 8
```

Размер входящего трафика RDMA для восстановления кэша XON.

```
mmu xon-threshold-offset pg 5 set 12
```

Настройка гарантированного кэша приоритетной записи RDMA.



```
mmu pg-guarantee pg 5 set 17
```

```
!
```

Включение PFC на интерфейсе на основе PG (трафик носителя RDMA)

Включение взаимоблокировки PFC на основе PG (трафик носителя RDMA) на интерфейсе.

```
interface AggregatePort 1
```

```
priority-flow-control nodrop 5 on
```

```
priority-flow-control deadlock cos-value 5 enable
```

```
!
```

3.3. Параметры MMU на Uplink интерфейсе и включение PFC (если планирование RDMA трафика производится только на данном устройстве, следующая конфигурация не требуется)

```
interface Hundred Gigabit Ethernet 0/55
```

```
priority-flow-control nodrop 5 on
```

```
priority-flow-control deadlock cos-value 5 enable
```

```
mmu queue-threshold output unicast 5 set 100
```

```
mmu queue-threshold output unicast 6 set 100
```

```
mmu headroom-threshold pg 5 set 491
```

```
mmu xoff-threshold pg 5 set 8
```

```
mmu xon-threshold-offset pg 5 set 12
```

```
mmu pg-guarantee pg 5 set 17
```

```
!
```

Настройка размера headroom

3.4. Глобальное включение PFC

Отключение функции DCB, чтобы избежать влияния журнала, связанного с согласованием с сервером.

```
no dcbx enable
```

Время интервала статистики PFC

```
priority-flow-control counting-interval 1
```

Настройка параметра взаимоблокировки PFC

```
priority-flow-control deadlock cos-value 5 detect 10 recover 100
```

```
priority-flow-control deadlock limit frequency 3 period-second 20
```

```
priority-flow-control deadlock limit cos-value 5 enable
```

```
!
```

4. ECN

4.1. Настройка ECN на downlink интерфейсе



Конфигурация QOS производится на логическом интерфейсе.

```
interface AggregatePort 1
```

Включение ECN для RDMA Traffic Queue

```
qos queue 5ecn
```

```
qos queue 6ecn
```

Каждая очередь устанавливает минимальный и максимальный порог отбрасывания WRED. Для очереди RDMA с включенным ECN порог минимальной отметки ECN и максимальный порог отметки ECN не будут отброшены.

```
wrr-queue random-detect min-threshold cell 0 2500 2500
```

```
wrr-queue random-detect min-threshold cell 1 2500 2500
```

```
wrr-queue random-detect min-threshold cell 2 2500 2500
```

```
wrr-queue random-detect min-threshold cell 3 2500 2500
```

```
wrr-queue random-detect min-threshold cell 4 2500 2500
```

```
wrr-queue random-detect min-threshold cell 5 400 400
```

```
wrr-queue random-detect min-threshold cell 6 2500 2500
```

```
wrr-queue random-detect max-threshold cell 0 10000 10000
```

```
wrr-queue random-detect max-threshold cell 1 10000 10000
```

```
wrr-queue random-detect max-threshold cell 2 10000 10000
```

```
wrr-queue random-detect max-threshold cell 3 10000 10000
```

```
wrr-queue random-detect max-threshold cell 4 10000 10000
```

```
wrr-queue random-detect max-threshold cell 5 1625 1625
```

```
wrr-queue random-detect max-threshold cell 6 10000 10000
```

Максимальная вероятность отброса WRED и максимальная вероятность отметки ECN

```
wrr-queue random-detect probability 0 20 20
```

```
wrr-queue random-detect probability 1 20 20
```

```
wrr-queue random-detect probability 2 20 20
```

```
wrr-queue random-detect probability 3 20 20
```

```
wrr-queue random-detect probability 4 20 20
```

```
wrr-queue random-detect probability 5 20 20
```

```
wrr-queue random-detect probability 6 20 20
```

!

4.2. Настройка ECN на uplink интерфейсе

Разница заключается в пороге WRED.

```
interface Hundred GigabitEthernet 0/56
```

```
qos queue 5ecn
```

```
qos queue 6ecn
```

```
wrr-queue random-detect min-threshold cell 0 3000 3000
```



wrr-queue random-detect min-threshold cell 1 3000 3000
wrr-queue random-detect min-threshold cell 2 3000 3000
wrr-queue random-detect min-threshold cell 3 3000 3000
wrr-queue random-detect min-threshold cell 4 3000 3000
wrr-queue random-detect min-threshold cell 5 1000 1000
wrr-queue random-detect min-threshold cell 6 3000 3000
wrr-queue random-detect max-threshold cell 0 15000 15000
wrr-queue random-detect max-threshold cell 1 15000 15000
wrr-queue random-detect max-threshold cell 2 15000 15000
wrr-queue random-detect max-threshold cell 3 15000 15000
wrr-queue random-detect max-threshold cell 4 15000 15000
wrr-queue random-detect max-threshold cell 5 2000 2000
wrr-queue random-detect max-threshold cell 6 15000 15000
wrr-queue random-detect probability 0 20 20
wrr-queue random-detect probability 1 20 20
wrr-queue random-detect probability 2 20 20
wrr-queue random-detect probability 3 20 20
wrr-queue random-detect probability 4 20 20
wrr-queue random-detect probability 5 20 20
wrr-queue random-detect probability 6 20 20
!



5. ОБЩАЯ ИНФОРМАЦИЯ

5.1. Замечания и предложения

Мы всегда стремимся улучшить нашу документацию и помочь вам работать лучше, поэтому мы хотим услышать вас. Мы всегда рады обратной связи, в особенности:

- ошибки в содержании, непонятные или противоречащие места в тексте;
- идеи по улучшению документации, чтобы находить информацию быстрее;
- неработающие ссылки и замечания к навигации по документу.

Если вы хотите написать нам по поводу данного документа, то используйте, пожалуйста, форму обратной связи на qtech.ru.

5.2. Гарантия и сервис

Процедура и необходимые действия по вопросам гарантии описаны на сайте QTECH в разделе «Поддержка» -> «[Гарантийное обслуживание](#)».

Ознакомиться с информацией по вопросам тестирования оборудования можно на сайте QTECH в разделе «Поддержка» -> «[Взять оборудование на тест](#)».

Вы можете написать напрямую в службу сервиса по электронной почте sc@qtech.ru.

5.3. Техническая поддержка

Если вам необходимо содействие в вопросах, касающихся нашего оборудования, то можете воспользоваться нашей автоматизированной системой запросов технического сервис-центра helpdesk.qtech.ru.