

Índice

[Introdução](#)

[Antes de Começar](#)

[Convenções](#)

[Pré-requisitos](#)

[Componentes Utilizados](#)

[Enfileiramento](#)

[Enfileiramento de prioridade estrita/enfileiramento de latência baixa](#)

[Compartilhando largura de banda de link](#)

[Modelagem de tráfego](#)

[Programação de fila de trânsito](#)

[Informações Relacionadas](#)

[Introdução](#)

O Catalyst 4000 com Supervisor III (WS-X4014) ou Supervisor IV (WS-X4515) suporta recursos avançados de QoS (Qualidade de Serviço), incluindo classificação, vigilância, marcação, enfileiramento e programação. Este documento trata dos recursos de enfileiramento e programação, incluindo a forma do tráfego, compartilhamento e enfileiramento de estrita prioridade/latência baixa. O enfileiramento determina a forma como os pacotes são colocados em várias filas na interface de entrada e a programação determina a forma como (em momentos de congestionamento) se dá prioridade ao tráfego de alta prioridade em relação ao tráfego de baixa prioridade.

[Antes de Começar](#)

[Convenções](#)

Para obter mais informações sobre convenções de documento, consulte as [Convenções de dicas técnicas Cisco](#).

[Pré-requisitos](#)

Os leitores deste documento devem estar cientes da seguinte informação:

- A prioridade da camada 2 (L2) dos quadros é baseada em um valor do Classe de serviço (CoS), que esteja disponível no encabeçamento do InterSwitch Link (ISL) (três mordidas o mais menos significativas no campo do usuário 4-bit) e no encabeçamento do 802.1Q (três bit mais significativo em 2-byte etiquetam o campo de informação de controle).
- A priorização de pacotes de camada 3 (L3) é baseada no valor de DSCP (Differentiated Services Code Point), que está disponível no byte de ToS (Tipo de serviço) no cabeçalho de IP (os seis bits mais significativos) ou no valor de precedência de IP do byte de ToS (os três bits mais significativos).

- [Consulte o manual de configuração de software para obter assistência de configuração adicional.](#)

Componentes Utilizados

As informações neste documento estão baseadas nas seguintes versões de software em um Supervisor III (WS-X4014):

- Software Release 12.1(8)EW de Cisco IOS®

Nota: O Supervisor IV é apoiado primeiramente no Cisco IOS Software Release 12.1(12c)EW. As características descritas neste documento aplicam-se ao Supervisor IV também, a menos que diferenciadas de outra maneira explicitamente.

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. Todos os dispositivos utilizados neste documento foram iniciados com uma configuração (padrão) inicial. Se você estiver trabalhando em uma rede ativa, certifique-se de que entende o impacto potencial de qualquer comando antes de utilizá-lo.

Enfileiramento

O Catalyst 4000 Supervisor III e IV usam uma arquitetura de switching de memória compartilhada e podem fornecer recursos de enfileiramento e programação para as placas de linha existentes. Como o Supervisor fornece a arquitetura de switching sem bloqueio, não há filas de entrada. Os pacotes são encaminhados pelo painel traseiro para a porta de saída ou egressa. O lado de saída da interface fornece quatro filas de transmissão. O tamanho da fila está atualmente fixo em 240 pacotes para portas FastEthernet e 1920 pacotes para interfaces Gigabit Ethernet de não bloqueio. O não-bloqueio significa que as portas não estão com excesso de assinaturas na conexão com o painel traseiro. Uma lista de portas Gigabit Ethernet sem bloqueio é apresentada a seguir:

- portas de uplink no Supervisor Engine III (WS-X4014) e IV (WS-X4515)
- portas na placa WS-X4306-GB
- duas portas 1000BASE-X na placa de linha WS-X4232-GB-RJ
- primeiras duas portas na placa de linha WS-X4418-GB
- duas portas 1000BASE-X na placa de linha WS-X4412-2GB-TX

Obstruir (sobre-subscrita) o tamanho da fila da porta de Ethernet Gigabit é fixada atualmente em 240 pacotes também. As portas de bloqueio estão listadas a seguir:

- Portas 10/100/1000 de T na placa de linha WS-X4412-2GB-TX
- portas na placa de linha WS-4418-GB, exceto as duas primeiras portas
- portas na placa de linha WS-X4424-GB-RJ45
- portas na placa de linha WS-X4448-GB-LX.
- portas na placa de linha WS-X4448-GB-RJ45

Nota: O tamanho da fila é baseado no número de pacotes e não no tamanho dos pacotes. Atualmente, o Supervisor III não apoia nenhum mecanismo de fuga de congestionamento tal como o Weighted Random Early Detection (WRED) para os transmitir fila.

Nota: O Supervisor IV apoia a característica ativa do gerenciamento de fila (AQM) no Cisco IOS Release 12.1(13)EW e Mais Recente. O AQM é uma técnica de fuga de congestionamento que

atue antes que o excesso de buffer ocorra. O AQM é conseguido com da limitação dinâmica do buffer (DBL). O DBL segue o comprimento da fila para cada fluxo de tráfego no interruptor. Quando o comprimento da fila de um fluxo específico excede seu limite, o DBL deixará cair pacotes ou ajustará os bits da notificação de congestionamento explícito (ECN) nos cabeçalhos de pacote de informação. Para obter mais informações sobre de como configurar o DBL, refira [configurar QoS](#).

Quando o QoS é desativado, os pacotes são confiáveis para o DSCP recebido nos pontos de entrada e colocados nas filas apropriadas. Essas filas usam o método round-robin.

Quando o QoS estiver habilitado, os pacotes são enfileirados com base no DSCP interno, que é derivado do CoS/DSCP de entrada, utilizando estados confiáveis de porta, ou da configuração padrão de CoS/DSCP na porta de entrada ou na marcação baseada em Lista de acesso (ACL)/Classe. A fila é selecionada com base no mapeamento dscp-tx-queue global, que é inteiramente configurável. O mapeamento pode ser indicado como segue:

```
Switch#show qos maps dscp tx-queueDSCP-TxQueue Mapping Table (dscp = d1d2)d1 : d2 0 1 2 3 4
5 6 7 8 9-----0 : 01 01 01 01 01 01 01 01 01 01 011 : 01
01 01 01 01 01 02 02 02 022 : 02 02 02 02 02 02 02 02 023 : 02 02 03 03 03 03 03 03
034 : 03 03 03 03 03 03 03 03 04 045 : 04 04 04 04 04 04 04 04 046 : 04 04 04 04
```

O mapeamento acima é o padrão. Se necessário, traçar pode ser mudado emitindo o **comando qos map dscp dscp-values to tx-queue queue-id**. Por exemplo, para mapear um valor DSCP de 50 para tx-queue 2, a configuração a seguir é feita no modo de configuração global:

```
Switch(config)#qos map dscp 50 to tx-queue 2!--- You can verify to make sure the changes have
been made. Switch #show qos maps dscp tx-queueDSCP-TxQueue Mapping Table (dscp = d1d2)d1 : d2 0
1 2 3 4 5 6 7 8 9-----0 : 01 01 01 01 01 01 01 01 01 01
011 : 01 01 01 01 01 01 02 02 02 022 : 02 02 02 02 02 02 02 02 023 : 02 02 03 03 03
03 03 03 03 034 : 03 03 03 03 03 03 03 03 04 045 : 02 04 04 04 04 04 04 04 046 : 04
04 04 04
```

Para mais informações sobre os passos de configuração para alterar o mapeamento, consulte o seguinte documento:

- [Guia de configuração](#)

Devido comutar a limitação dos circuitos integrados do aplicativo específicos (ASIC), se a porta de ingresso é ajustada ao Trust-cos, transmitir CoS é igual ao pacote recebido CoS ou ao padrão CoS (para pacotes sem etiqueta) configurado na porta. Se uma política for configurada para definir o DSCP para o pacote emitindo o comando set ip dscp value para tais pacotes, eles serão usados como a origem do DSCP interno em vez do CoS padrão/ de pacote e serão enfileirados nas filas apropriadas. Se a porta não for confiável para CoS, o CoS de saída será baseado no valor do DSCP interno.

[Enfileiramento de prioridade estrita/enfileiramento de latência baixa](#)

Se necessário, a fila de transmissão 3 pode ser configurada como fila de prioridade máxima, para que os pacotes nela enfileirados sejam agendados para serem transmitidos antes dos pacotes enfileirados nas outras filas, desde que não excedam o valor de compartilhamento configurado. Isso está explicado na seguinte seção.

A característica da prioridade estrita é desabilitada à revelia. O mapeamento padrão enfileiraria pacotes com CoS 4 e 5 e o DSCP 32 com 47 no transmitir fila 3. DSCP ao traço do Tx-queue podem ser alterados como desejado de modo que os pacotes desejados sejam enfileirados na fila de alta prioridade.

Para não morrer de fome os pacotes de prioridade baixa, esta fila precisa de ser configurada primeiramente para o volume baixo, mas o tráfego de alta prioridade, tal como o tráfego de voz, e não para o tráfego maioria da prioridade baixa TCP/IP. Igualmente recomenda-se configurar dar forma/que compartilha para a fila de alta prioridade se uma precisa de impedir a inanição das outras filas de prioridade não-estritas. Ao configurar a modelagem/compartilhamento, os outros pacotes de baixa prioridade serão agendados, uma vez que o valor de molde/compartilhamento da fila estrita não foi atendido.

```
Switch#show run interface gigabitEthernet 1/1interface GigabitEthernet1/1 no switchport ip address 10.1.1.1 255.255.255.0 tx-queue 3 priority high end
```

Compartilhando largura de banda de link

Catalyst 4000 Supervisor III e apoio IV o comando **bandwidth**, que é um secundário-comando sob o comando **tx-queue**. Este comando permite uma largura de banda mínima garantida a cada um dos quatro transmitir fila. Este comando não deve ser confundido com o comando **bandwidth** do nível de interface que é usado para finalidades do protocolo de roteamento. Isso, junto com o mapeamento da fila TX de DSCP, oferece o controle granular de quanta largura de banda é garantida para cada classe de tráfego enfileirado em cada uma das quatro filas. Tipicamente, o tráfego de alta prioridade tal como o tráfego de voz é garantido uma determinada quantidade mínima de tráfego durante épocas da congestão com o Enfileiramento de prioridade estrita, com uma parte configurada para a partilha do transmitir fila 3. da largura de banda de enlace é apoiado somente nas portas Ethernet de gigabit não bloqueada. Este recurso atualmente não está disponível em portas de Ethernet Gigabits de bloqueio ou interfaces Fast Ethernet 10/100.

Quando QoS é permitido globalmente no interruptor, todas as quatro filas estão atribuídas à revelia uma largura de banda mínima do 250 Mbps em todas as portas. Pode ser necessário alterar as configurações padrão para ter certeza de que elas correspondam às configurações desejadas para o aplicativo ou para a rede em questão.

```
Switch#show run interface gigabitEthernet 1/1interface GigabitEthernet1/1 no switchport ip address 10.1.1.1 255.255.255.0 tx-queue 1 bandwidth 500 mbps tx-queue 2 bandwidth 25 mbps tx-queue 3 bandwidth 50 mbps priority high tx-queue 4 bandwidth 200 mbpsendSwitch#show qos interface GigabitEthernet 1/1QoS is enabled globallyPort QoS is enabledPort Trust State: 'untrusted'Default DSCP: 0 Default CoS: 0tx-Queue Bandwidth ShapeRate Priority QueueSize (bps) (BPS) (packets) 1 500000000 disabled N/A 1920 2 250000000 disabled N/A 1920 3 50000000 disabled high 1920 4 200000000 disabledN/A 1920
```

O interruptor atualmente não valida que a soma da parte da largura de banda pelo 1 Gbps do <= da fila. Por exemplo, se Q1 = 300 Mbps, Q2 = 200 Mbps, Q3= 100 Mbps e Q4 = 500 Mbps, estamos excedendo a largura de banda total disponível para essa interface em 1 Gbps. Para compreender como o interruptor se comportaria neste cenário de assinatura em excesso, nós precisamos uma compreensão de como a programação trabalha.

Quando a taxa de saída de uma fila de transmissão está abaixo dos valores configurados de compartilhamento e forma, ela é considerada uma fila de alta prioridade. Inicialmente, todas as filas serão de alta prioridade, já que a nenhuma delas recebeu sua parte, portanto, serão servidas em rodízio (observe que uma fila configurada como de alta prioridade sempre será servida primeiro se não estiver vazia até encontrar sua parte). Como algumas das filas atendem ao compartilhamento, se houver qualquer outra fila com prioridade alta, ela será atendida. Se não houver nenhuma fila de prioridade alta, todas as filas de prioridade baixa (que já encontraram seu compartilhamento) serão atendidas no estilo round-robin.

Com base na descrição da operação acima, em nosso cenário de exemplo, o Q1, Q2 e Q3 conseguirão suas partes, mas não o Q4 em momentos de congestionamento, pois a interface não

pode alocar uma largura de banda acima da sua largura de banda física disponível. Deve-se tomar cuidado na escolha dos valores compartilhados de acordo com os requisitos do usuário/aplicativo.

Modelagem de tráfego

O Catalyst 4000 Supervisor III e o IV suportam outros recursos de modelagem de tráfego além do recurso de vigilância. Os recursos de modelagem podem ser configurados por fila de transmissão no FastEthernet bem como no Gigabit Ethernet. A modelagem limite a largura de banda transmitida por fila por segundo ao valor máximo configurável de 16 Kbps a 1 Gbps (100 Mbps para a porta FastEthernet). Dar forma tem a variação muito baixa do valor configurado como a decisão para transmitir um pacote de uma fila específica é feito pelo pacote.

```
Switch#show run interface FastEthernet 5/9interface FastEthernet5/9 no switchport no snmp trap
link-status ip address 10.1.1.1 255.255.255.0 tx-queue 1  shape 50 mbps tx-queue 2  shape 35
mbps tx-queue 3  priority high  shape 5 mbps tx-queue 4  shape 10 mbps  Switch#show qos
interface FastEthernet 5/9QoS is enabled globallyPort QoS is enabledPort Trust State:
'untrusted'Default DSCP: 0 Default CoS: 0tx-Queue Bandwidth ShapeRate Priority QueueSize (BPS)
(BPS) (packets) 1 N/A 50000000 N/A 240 2 N/A 35000000 N/A 240 3 N/A
5000000 high 240 4 N/A 10000000 N/A 240
```

Programação de fila de trânsito

Os pacotes são enfileirados com base no DSCP interno em uma das quatro filas descritas mais cedo. O DSCP interno pode ser derivado do ingresso DSCP, da porta de ingresso DSCP, ou do Class-based Marking. A programação do transmitir fila acontece como segue. Se dando forma é configurado, o pacote no transmitir fila está verificado se esteja dentro da máxima configurada do valor da forma. Se exceder o valor, será enfileirado e não será transmitido.

Se o pacote está qualificado, o recurso sharing/strict priority (prioridade de compartilhamento/estrita) é considerado. Primeiro, a preferência é concedida aos pacotes enfileirados com prioridade estrita, desde que estejam abaixo do parâmetro de modelagem configurado para a fila. Depois que a fila de prioridade estrita está prestada serviços de manutenção (isto é, nenhum pacote na fila de prioridade estrita ou nela encontrou sua parte), os pacotes enfileirados na fila de prioridade não-estrita estão prestados serviços de manutenção no arredondamento robin. Desde que há três tais filas, compartilhar configurada para aquelas filas é considerada outra vez. Por exemplo, se o transmitir fila 1 não encontrou sua parte, tem a prioridade mais alta do que o transmitir fila 2, que encontrou sua parte. Quando esses pacotes enfileirados de prioridade mais alta são desenfileirados, os pacotes das filas que já encontraram seu compartilhamento são considerados.

Nota: A prioridade mais alta neste contexto não significa o melhor DSCP, o CoS, ou o valor de precedência IP. Baseia-se unicamente no fato de uma determinada fila ter ou não encontrado seu compartilhamento. Se a fila de prioridade não-estrita particular não encontrou sua parte, considera-se uma fila mais prioritária entre a fila de prioridade não-estrita que encontrou sua parte.

Informações Relacionadas

- [Entendendo e configurando QoS](#)
- [Perguntas freqüentes sobre qualidade de serviço no Catalyst 4000](#)
- [A Vigilância e a Marcação de QoS com Catalyst 4000 Supervisor Engine 3](#)

- [Apoio de produtos de LAN](#)
- [Suporte de tecnologia de switching de LAN](#)
- [Suporte Técnico - Cisco Systems](#)