

# Entender a Alocação de Buffer de Fila nos Switches Catalyst 9000

## Contents

---

[Introdução](#)

[Pré-requisitos](#)

[Requisitos](#)

[Componentes Utilizados](#)

[Produtos Relacionados](#)

[Informações de Apoio](#)

[Terminologia](#)

[Buffers](#)

[Algoritmo de Alocação de Buffer de Fila de Switch](#)

[Configurar](#)

[Cenário 1: Determinar o Valor de Buffer Base para uma Determinada Velocidade de Porta](#)

[Cenário 2: Determine o buffer flexível e rígido para várias filas e taxas explícitas](#)

[Cenário 2.1 : Modificar Multiplicador Configurado pelo Usuário do Cenário 2](#)

[Cenário 3: Determine o buffer flexível e rígido para várias filas e uma proporção implícita](#)

[Cenário 3.1 : Determine o buffer flexível e rígido para várias filas e várias taxas implícitas](#)

[Cenário 4 : A soma da razão de buffers de fila entre as filas é menor que 100](#)

[Cenário 5 : Determine o buffer flexível e rígido de uma classe de prioridade nível 2](#)

[Limite da fila](#)

[Informações Relacionadas](#)

---

## Introdução

Este documento descreve como prever a alocação de buffer de fila para filas de tráfego nos Catalyst 9000 Series Switches.

## Pré-requisitos

### Requisitos

A Cisco recomenda que você tenha conhecimento destes tópicos:

- Conceitos e processos de Qualidade de Serviço (QoS - Quality of Service) de como os pacotes são Marcados, Enfileirados e Agendados
- Configuração de QoS Cisco MQC

## Componentes Utilizados

As informações neste documento são baseadas nestas versões de software e hardware:

- Cisco Catalyst 9500-25Y4C
- Cisco IOS® XE 17.4.1

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 a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

## Produtos Relacionados

Este documento também pode ser usado com as seguintes versões de hardware e software:

- Switches Catalyst 9200 - 9600 Series
- Catalyst 9300X e 9400X
- Todas as versões do software Cisco IOS XE



Note: Este documento não se aplica ao 9500X ou 9600X, pois eles usam uma arquitetura ASIC e QoS diferente.

---

## Informações de Apoio

Para obter uma visão geral técnica da QoS nos Catalyst 9000 Series Switches, consulte: [White paper sobre QoS e enfileiramento do Catalyst 9000](#).

Frequentemente, a alocação de buffer precisa ser ajustada como resposta a quedas de saída indesejadas para uma classe específica de tráfego. Mais informações sobre como diagnosticar e solucionar problemas de quedas de saída nos Catalyst 9000 Series Switches podem ser encontradas neste artigo: [Solucionar problemas de descartes de saída nos switches Catalyst 9000](#)

## Terminologia

qos	Qualidade de Serviço	Um conceito/grupo de recursos relacionados relacionados à classificação, marcação, fila e programação de tráfego de entrada e saída de um dispositivo de rede
DSCP	Differentiated Services Code Point	Um mecanismo de classificação de tráfego contido no cabeçalho IP de um pacote
CoS	Classe de serviço	Um mecanismo de classificação de tráfego contido no cabeçalho do quadro Ethernet de um pacote
ACE	Entrada de Controle de Acesso	Uma única regra ou linha em uma ACL (Access Control List, lista de controle de acesso)
ACL	Lista de Controle de Acesso	Um grupo de Entradas de Controle de Acesso (ACEs - Access Control Entries) usado por vários recursos para corresponder o tráfego e executar uma ação
ASIC	Circuito integrado específico da aplicação	Um chip de computador projetado para executar uma tarefa específica ou um conjunto de tarefas com alta eficiência.
UADP	Plano de dados do Unified Access	O Cisco ASIC usado nos Catalyst 9000 Series Switches para executar muitas tarefas de processamento de pacotes de rede.
PBC	Complexo de buffer de pacote	Subsistema Cisco UADP ASIC que serve como buffer de pacote central para processar, enfileirar e programar pacotes.
AQM	Gerenciamento de fila ativa	Subsistema Cisco UADP ASIC que gerencia a fila de tráfego e agenda ações para portas de rede.
DTS	Limite e escala dinâmicos	Tecnologia Cisco UADP ASIC que ajusta e dimensiona dinamicamente os buffers nas portas para otimizar a utilização do hardware

## Buffers

Como conceito, os buffers são memórias usadas para absorver rajadas transitórias de dados, quando os dados comutados ou roteados para uma porta excedem a capacidade das portas de colocar dados no fio. Uma porta tem uma taxa fixa na qual transmite e remove dados da fila. Um buffer, conceitualmente, é simplesmente um local para armazenar ou enfileirar dados até que sejam transmitidos para fora da interface.

Nos Catalyst 9000 Series Switches, a palavra buffer tem dois usos. O buffer do sistema como um todo é conhecido como o complexo de buffer de pacotes (PBC) do ASIC. A palavra buffers também pode se referir a uma pequena unidade do PBC. Um buffer é alocado para portas em uma base por fila. Em outras palavras, uma fila de porta recebe uma quantidade de pequenos buffers individuais do buffer geral do sistema.

Em plataformas baseadas no Cisco UADP ASIC, um buffer contém até 256 bytes de dados e os buffers são vinculados para representar quadros maiores que 256 bytes.

### Algoritmo de Alocação de Buffer de Fila de Switch

- Uma porta tem um valor de buffer base codificado alocado que pode ser dimensionado pelo usuário ou pelo sistema (mas o número base não pode ser alterado, o que é importante para entender como prever resultados)
- O valor de buffer base para um determinado tipo de porta depende de uma combinação específica de hardware (modelo de switch), versão de software e velocidade de porta. Por exemplo, uma porta Gigabit Ethernet no Catalyst C9500-25Y4C começa com um valor de buffer base diferente de uma porta Gigabit Ethernet no Catalyst 9200L
- Esse valor base é dividido em N números de filas - com base na configuração.

O cálculo final do buffer disponível por fila é influenciado por estes fatores:

- Número de filas configuradas
- A relação fila-buffers, um comando que permite ajustar como o buffer base é dividido entre as filas de porta configuradas.
- Um multiplicador de buffer de base configurado pelo usuário (configurado como `qos queue-softmax-multiplier <percent>`)
- Um multiplicador de buffer de base definido pelo sistema (não configurável pelo usuário, 4x ou 400%) - enquanto isso é aplicado por padrão, há algumas circunstâncias em que ele é removido; portanto, é importante entender que esse valor 4x existe para prever o algoritmo de alocação

- Ambos os multiplicadores afetam apenas o que é chamado de buffers soft

Os buffers suaves são buffers compartilhados entre portas. Esses buffers são chamados de soft porque não são garantidos para a porta.

O sistema aloca os buffers suaves intencionalmente. Isso permite que qualquer porta use um grande número de buffers, se necessário, mas à medida que mais portas precisarem de buffers, todas as portas e filas serão reduzidas de forma dinâmica e justa como parte do processo Cisco UADP ASIC DTS.

Em resumo, os buffers suaves, chamados de softmax nas saídas, são um valor máximo oportunista. Uma porta usa apenas o softmax completo se essa quantidade de buffer estiver disponível no buffer geral do sistema. À medida que a demanda de buffer aumenta entre outras portas e filas, o buffer máximo disponível para a porta é menor.

Os buffers rígidos são buffers explicitamente reservados para uma porta e não são afetados pelo processo DTS. Como os buffers rígidos são buffers garantidos, a soma do total de buffers rígidos alocados para as portas nunca excede o segmento PBC dedicado a esses buffers rígidos.

Os mecanismos que controlam a escala ativa do buffer soft são conhecidos como DTS (Dynamic Threshold and Scale), descritos no [white paper sobre QoS da série Catalyst 9000](#).

O tamanho dos segmentos de PBC dedicados aos buffers de hardware e software da fila muda dinamicamente à medida que você configura o sistema e pode ser visto como AQM GlobalSoftLimit e GlobalHardLimit nesta saída:

```
<#root>
```

```
C9500#
```

```
show platform hardware fed active qos queue stats interface twe1/0/1
```

```
-----  
AQM Global counters  
GlobalHardLimit:
```

```
18072
```

```
| GlobalHardBufCount: 0  
GlobalSoftLimit:
```

```
37224
```

```
| GlobalSoftBufCount: 0
```

C9500#

show platform hardware fed active qos queue config interface twl/0/1

Asic:0 Core:1 DATA Port:20 GPN:101 LinkSpeed:0x12  
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 160 - 167  
DrainFast:Disabled PortSoftStart:2 - 4320 BufferSharing:Disabled  
DTS

Hardmax

Softmax

PortSMin	GlbISMin	PortStEnd	QEnable
0	1	2	
480			
	3		
1920			
16	960	0	0
		4	5760
			En

<--- default configuration has a mix of hard buffer and soft buffer in queue 0

1	1	0
0		

	4	
2880		

16	1440	8	720	4	5760	En
----	------	---	-----	---	------	----

<--- default configuration has two queues so some buffers are seen in queue 1

C9500(config)#

```
policy-map test
```

```
C9500(config-pmap)#
```

```
class class-default
```

```
C9500(config-pmap-c)#
```

```
priority level 1
```

```
<--- Priority level 1 queue configuration on first queue, which is queue 0 in the next output
```

```
C9500(config-pmap-c)#
```

```
exit
```

```
C9500(config-pmap)#
```

```
exit
```

```
C9500(config)#
```

```
int tw1/0/1
```

```
C9500(config-if)#
```

```
service-policy output test
```

```
C9500(config-if)#
```

```
end
```

```
C9500#
```

```
show platform hardware fed active qos queue config interface twel/0/1
```

```
Asic:0 Core:1 DATA Port:20 GPN:101 LinkSpeed:0x12  
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 160 - 167  
DrainFast:Disabled PortSoftStart:4 - 1800 BufferSharing:Disabled  
DTS
```

Hardmax

Softmax

```
PortSMin GblSMin PortStEnd QEnable
-----
0    1  4
1200
7
1200
0    0 0    0 3  2400  En
```

<--- Hardmax increased to 1200 from 480 in queue 0, softmax reduced to 1200 from 1920

```
1    1 0    0 0    0 0    0 0    0 3  2400  En
```

<--- queue 1 now no longer has any values, as no second queue is configured

C9500#

show platform hardware fed active qos queue stats interface twel/0/1

```
-----
AQM Global counters
GlobalHardLimit:
```

18792

| GlobalHardBufCount: 0

<--- GlobalHardLimit increased to 18792 from 18072, or by 720

GlobalSoftLimit:

36504

| GlobalSoftBufCount: 0

<---

GlobalSoftLimit decreased from 37224 to 36504, or by 720



Note: Observe o aumento em GlobalHardLimit e a diminuição proporcional em GlobalSoftLimit.

---

Além disso, quando você configura o nível de prioridade 1, o softmax dessa fila é definido estaticamente para ser exatamente igual ao hardmax. Você só pode modificar o buffer rígido de uma fila de nível de prioridade 1.

A alteração em GlobalHardLimit e GlobalSoftLimit é igual a 720. Isso também é igual à alteração em hardmax após a configuração.

Os cenários neste documento explicam como computar e prever alocações de softmax e hardmax em várias configurações de mapa de políticas.

## Configurar

### Cenário 1: Determinar o Valor de Buffer Base para uma Determinada Velocidade de Porta

Um valor de buffer final de filas é em parte uma função de um valor base que é alocado primeiro entre as filas. Isso é então posteriormente multiplicado no caso de buffers soft.

Os fatores de multiplicação, em combinação com outros comportamentos implícitos, tornam a determinação de um valor final para uma determinada fila, com uma determinada configuração, um desafio.

A primeira etapa para esclarecer a alocação de buffer de fila resultante é determinar o valor de buffer base.

Para fazer isso, utilize uma fila de prioridade, que recebe o buffer rígido diretamente proporcional ao número de filas ou à proporção de buffers de fila configurada.

Com uma configuração específica, você pode derivar explicitamente a quantidade de buffer base alocada para uma determinada velocidade de porta.

Configurar e atribuir todos os buffers a uma única fila não multiplicada (uma fila de nível de prioridade 1)

Neste exemplo, a classe class-default é usada para corresponder todo o tráfego, pois nenhuma outra classe está configurada.

```
<#root>
```

```
Switch(config)#
```

```
policy-map test1
```

```
Switch(config-pmap)#
```

```
class class-default
```

```
Switch(config-pmap-c)#
```

```
priority level 1
```

```
<--- Assign hard buffer to the port, which is not affected by multipliers
```

```
Switch(config-pmap-c)#
```

```
queue-buffers ratio 100
```

```
<--- Assign all buffers to this queue only
```

A configuração no exemplo anterior executa estas ações:

- O nível de prioridade 1 instrui o sistema a executar uma programação rigorosa para essa fila/classe e alocar buffer rígido para essa fila/classe.
  - O modo como os pacotes são agendados está além do escopo deste documento, mas uma fila estrita geralmente é atendida antes de todas as outras.

- A razão de fila-buffers 100 aloca 100/100 ou 100% do buffer base disponível para essa fila/classe.

Em um mapa de políticas com mais de uma classe, você não pode alocar 100% do buffer para uma única classe. Você deve alocar no mínimo 1/100 ou 1% para qualquer classe.

Em uma política com apenas uma classe, você tem apenas uma classe e pode alocar todo o buffer para ela.

Como observado anteriormente, uma fila de prioridade obtém buffers rígidos iguais à sua distribuição de buffer base de acordo com a razão de fila-buffers configurada. Um buffer rígido não está sujeito a nenhum multiplicador.

O buffer rígido é observado em saídas sob uma coluna intitulada Hardmax.

Agora você tem uma única classe de tráfego com buffers não sujeitos a nenhum multiplicador. Com isso, você pode derivar explicitamente a alocação de buffer base para essa velocidade de porta (e somente essa velocidade de porta nessa plataforma, outras diferem), porque o buffer base e o hardmax são iguais.

Buffer base = ?

Proporção da Fila 1 =  $100/100 = 1$

Hardmax para esta fila = Buffer base x taxa de fila 1

$$X = S \times 1$$

$$X / 1 = Y$$

$$X = Y$$

$X = Y = \text{Hardmax} = \text{buffer base} = 1200$  (consulte o Exemplo 2).

Neste exemplo, o mapa de políticas test1 é aplicado a uma interface como uma política de serviço de saída

<#root>

9500H(config)#

```
int tw1/0/3
```

```
9500H(config-if)#
```

```
service-policy output test1 <--- service policy that assigns all buffer to the first
```

```
9500H#
```

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12  
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183  
DrainFast:Disabled PortSoftStart:4 - 1800 BufferSharing:Disabled  
DTS
```

```
Hardmax
```

Softmax	PortSMin	GlbSMin	PortStEnd	QEnable
0	1	4		

```
1200
```

7	1200	0	0	0	0	3	2400	En
---	------	---	---	---	---	---	------	----

```
<--- hardmax 1200 - the maximum amount of buffer this port can use without multiplication
```

1	1	0	0	0	0	0	0	0	3	2400	En
2	1	0	0	0	0	0	0	0	3	2400	En
3	1	0	0	0	0	0	0	0	3	2400	En
4	1	0	0	0	0	0	0	0	3	2400	En
5	1	0	0	0	0	0	0	0	3	2400	En
6	1	0	0	0	0	0	0	0	3	2400	En
7	1	0	0	0	0	0	0	0	3	2400	En

```
<snip>
```

Como mostrado, o hardmax dessa fila de prioridade com 100% de buffer alocado é 1200.

Como hardmax é um valor totalmente não multiplicado / não escalado, e 100% do buffer é configurado para essa fila, a alocação de buffer base para esse modelo específico de switch, versão de software e velocidade de porta específica é 1200.

Outras velocidades de porta neste mesmo switch, e outros modelos de switches para a mesma velocidade de porta, recebem alocações de buffer base diferentes. Esta atribuição de base não é configurável pelo utilizador e deve ser derivada através de observação.

Outros cenários neste documento começam a usar o mesmo switch, software e velocidade de porta. Assim, todos eles assumem uma alocação base de 1200 para cálculos para determinar a alocação final de buffer.

---



Note: Observe que o softmax no exemplo anterior também é 1200.

---

Por design, uma fila de nível de prioridade 1 tem softmax igual exatamente a seu hardmax. Isso é intencional e não configurável pelo usuário.

Além disso, esse caso específico de alocação de softmax não é afetado pelos multiplicadores de softmax mostrados posteriormente. Somente uma fila de nível de prioridade 1 tem esse comportamento para softmax, que é o desejado.

## Cenário 2: Determine o buffer flexível e rígido para várias filas e taxas explícitas

Neste cenário, uma fila adicional é adicionada. Essa fila não usa o nível de prioridade 1 e, portanto, o softmax é dimensionado com multiplicadores.

Um multiplicador é configurado pelo usuário e o outro é oculto/não configurável.

Combine esses multiplicadores com o buffer base derivado para essa porta, nesse caso 1200, de acordo com o Cenário 1.

Algoritmicamente:

Proporção de fila atual = proporção fila-buffers para a fila / classe a ser prevista

Multiplicador Oculto = 400%

Multiplicador de Usuário = Valor percentual configurado em qos queue-softmax-multiplier <percent>. O padrão é 100%

$$\text{Softmax} = (\text{Buffer Base} \times (\text{Taxa de Fila Atual} / 100)) \times \text{Multiplicador Oculto} \times (\text{Multiplicador do Usuário} / 100)$$

<#root>

```
9500H(config)#
```

```
policy-map test2
```

```
9500H(config-pmap)#
```

```
class class1
```

```
9500H(config-pmap-c)#
```

```
priority level 1
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 50
```

```
<-- class 1 / first queue gets 50% of base buffer
```

```
9500H(config-pmap-c)#
```

```
class class-default
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 100 <-- required configuration due to priority queue, can be ignored for thi
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 50
```

```
<-- class 2 / first queue gets 50% of base buffer
```

Resumo dos valores:

- Buffer base = 1200 - Como encontrar esse número é mostrado no Cenário 1
- Multiplicador de Usuário = 100% - Valor padrão, que ainda não foi modificado
- Multiplicador Oculto = 4 - padrão e não configurável pelo usuário
- Taxa de fila atual (class1) = 50% - Configurada no mapa de política com razão fila-buffers

50

- Taxa de fila atual (padrão de classe) = 50% - Configurada no mapa de política com taxa de fila-buffers 50

Determine a alocação de buffer de Classe 1:

Como a classe1 é uma fila de prioridade, ela recebe hardmax (buffer rígido) e um caso especial de softmax não afetado pelos multiplicadores.

Hardmax Classe1 = (Buffer base x Taxa de fila atual (class1) / 100)

Hardmax Classe1 = 1200 x (50/100) = 600 - devido ao caso especial de uma fila de prioridade, pare todos os cálculos, atribua o resultado ao hardmax. Softmax é igual a Hardmax como regra para o nível de prioridade 1.

Determine a alocação de buffer de classe padrão:

Classe padrão = (Buffer base x (Taxa de fila atual (padrão de classe / 100)) x Multiplicador oculto x (Multiplicador do usuário / 100)

Classe class-default = [

[buffer base] 1200 x [Taxa de Fila Atual] (50/100) = 600

[resultado anterior] 600 x [Multiplicador oculto] 4 x [Multiplicador do Usuário] (100/100) = 2400

]

<#root>

9500H(config)#

int tw1/0/3

9500H(config-if)#

service-policy output test2

<-- apply the policy

9500H#

```
show platform hardware fed active qos queue config interface twl/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:5 - 3600 BufferSharing:Disabled
  DTS Hardmax Softmax PortSMin GlbSMin PortStEnd QEnable
  ---- -
0 1 5
```

600

9

600

```
0 0 0 0 1 4800 En
```

<-- Hardmax is 600 as predicted, Softmax is set equal to Hardmax due to priority level 1

```
1 1 0 0 10
```

2400

```
16 1200 8 600 1 4800 En
```

<--

Softmax is 2400 as predicted

<snip>

Resultado final: Q0 - Hardmax: 600 Softmax: 600. T1 - Softmax: 2400

### Cenário 2.1 : Modificar Multiplicador Configurado pelo Usuário do Cenário 2

Esse cenário começa do mesmo modo que o Cenário 2, exceto que agora você configura qos queue-softmax-multiplier 1200.

Isso multiplica os buffers softmax na configuração atual por 1200% ou um fator de 12.

Resumo dos valores:

- Buffer base = 1200 - Como encontrar esse número é mostrado no Cenário 1
- Multiplicador de Usuário = 1200% - Este é o valor alterado neste cenário, através de qos queue-softmax-multiplier 1200
- Multiplicador Oculto = 4 - padrão e não configurável pelo usuário
- Taxa de fila atual (class1) = 50% - Configurada no mapa de política com razão fila-buffers 50
- Taxa de fila atual (padrão de classe) = 50% - Configurado no mapa de política com razão fila-buffers 50

Determine a alocação de buffer de Classe 1:

Como a classe1 é uma fila de prioridade, ela recebe hardmax (buffer rígido) e um caso especial de softmax não afetado pelos multiplicadores.

Hardmax Classe1 = (Buffer base x Taxa de fila atual (class1) / 100)

Hardmax Classe1 = 1200 x (50/100) = 600 - devido ao caso especial de uma fila de prioridade, pare todos os cálculos, atribua o resultado ao hardmax. Softmax é igual a Hardmax como regra para o nível de prioridade 1.

Determine a alocação de buffer de classe padrão:

Classe padrão = (Buffer base x (Taxa de fila atual (padrão de classe / 100)) x Multiplicador oculto x (Multiplicador do usuário / 100)

Classe class-default =[

[buffer base] 1200 x [Taxa de Fila Atual] (50/100) = 600

[resultado anterior] 600 x [Multiplicador oculto] 4 x [Multiplicador do Usuário] (1200/100) = 28800

]

Configure qos queue-softmax-multiplier 1200 e observe as alterações para softmax (softmax é um valor de buffer máximo para essa fila, dimensionado dinamicamente com base no uso geral de buffer atual):

<#root>

9500H(config)#

```
qos queue-softmax-multiplier 1200
```

9500H#

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:3 - 31500 BufferSharing:Disabled
  DTS   Hardmax   Softmax   PortSMin   G1b1SMin   PortStEnd   QEnable
-----
0    1    5
```

600

5

600

```
0    0    0    0    6 42000    En
```

<-- Queue 0 does not change as its configured with priority level 1

```
1    1    0    0    6
```

28800

```
1    900    1    900    6 42000    En
```

<-- Softmax increases by 12x to 28800 from 1200 due to queue-softmax-multiplier 1200

<snip>

Resultado final: Q1 - Hardmax: 600, Softmax 600. 2º trim. - Softmax: 28800

Cenário 3: Determine o buffer flexível e rígido para várias filas e uma proporção implícita

Neste cenário, cinco filas são configuradas, mas apenas quatro têm a razão de fila-buffers explicitamente definida. O buffer alocado para essas filas é o mesmo dos exemplos anteriores.

A fila não configurada recebe a diferença entre a soma de todos os buffers de fila configurados e 100.

Soma de taxas explicitamente configuradas = (taxa de buffer Q0) + (taxa de buffer Q1) ... (taxa de buffer final) - até 8 filas são suportadas nos Switches da série Catalyst 9000, portanto, você pode adicionar até 8 taxas

Taxa implícita restante = (100 - Soma de taxas explicitamente configuradas).

Taxa Implícita Restante é o valor que é atribuído a uma fila que não tem a taxa de buffers de fila configurada.

Mapa de políticas em uso para este cenário:

```
<#root>
```

```
9500H(config)#
```

```
policy-map test3
```

```
9500H(config-pmap)#
```

```
class class1
```

```
9500H(config-pmap-c)#
```

```
priority level 1
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 20
```

```
9500H(config-pmap-c)#
```

```
class class2
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10 <-- no queue-buffers ratio statement for this class
```

```
9500H(config-pmap-c)#
```

```
class class3
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 10
```

```
<-- rest of queues have an explicit queue-buffers ratio
```

```
9500H(config-pmap-c)#
```

```
class class4
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 10
```

```
9500H(config-pmap-c)#
```

```
class class-default
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 70
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 40
```

Resumo dos valores:

- Buffer base = 1200 - Como encontrar esse número é mostrado no Cenário 1
- Multiplicador de Usuário = 100% - Valor padrão, que ainda não foi modificado
- Multiplicador Oculto = 4 - padrão e não configurável pelo usuário
- Taxa de fila atual (class1) = 20% - Configurada no mapa de política com razão fila-buffers 20
- Taxa de fila atual (classe2) = Desconhecida, você deve resolver para esta - taxa de fila-buffers não está configurada para esta classe
- Taxa de fila atual (classe3) = 10% - Configurada no mapa de política com relação fila-buffers 10
- Proporção da fila atual (classe4) = 10% - Configurada no mapa de política com proporção de fila-buffers 10
- Taxa de fila atual (padrão de classe) = 40% - Configurado no mapa de política com razão fila-buffers 40

Calcular a taxa de buffer de fila que permanece:

Taxa implícita restante = (100 - Soma de taxas explicitamente configuradas).

$$100 - (20) - (10) - (10) - (50) = 20$$

$$\text{Taxa de fila atual (classe2)} = 20$$

Calcular alocação de buffer de fila final

Classe1 = [Buffer Base] 1200 x [Taxa de Fila Atual (classe1)] (20/100) = 240 - fila de prioridade, nenhum cálculo adicional

Classe2 = [

[Buffer Base] 1200 x [Taxa de Fila Atual(classe2)] (10/100) = 240 - alocação de buffer base para essa fila, mas ela deve ser multiplicada para obter o softmax para uma fila não prioritária

[Alocação de Buffer Base para esta fila] 120 x [Multiplicador oculto] 4 x [Multiplicador do Usuário] (100/100) = 960

]

Repita para as filas restantes:

Classe 3 = [

```
1200 x (10/100) = 120
120 x 4 x (100/100) = 480
]
```

Classe 4 = [

```
1200 x (10/100) = 120
120 x 4 x (100/100) = 480
]
```

Classe class-default = [

```
1200 x (40/100) = 480
600 x 4 x (100/100) = 1920
]
```

Resultado do mapa de políticas de test3 aplicado em comparação com a previsão:

```
<#root>
```

```
9500H(config)#
```

```
int tw1/0/3
```

```
9500H(config-if)#
```

```
service-policy output test3
```

```
9500H#
```

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:4 - 2880 BufferSharing:Disabled
  DTS   Hardmax   Softmax   PortSMin   G1b1SMin   PortStEnd   QEnable
  ----  -
0  1  5  240  8  240  0  0  0  0  6  3840  En
1  1  0   0  9
```

```
960
```

```
16  480  8  240  6  3840  En
```

<-- queue without queue buffers ratio configured receives any leftover ratio, as predicted

```
2  1  0    0 11  480 16  240  8  120  6 3840    En
3  1  0    0 11  480 16  240  8  120  6 3840    En
4  1  0    0  4 1920 16  960  8  480  6 3840    En
<snip>
```

Resultado final: Q0 - Hardmax: 240, Softmax 240. T1 - Softmax: 960, Q2 - Softmax: 480, Q3 - Softmax: 480, Q4 - Softmax: 480

Cenário 3.1 : Determine o buffer flexível e rígido para várias filas e várias taxas implícitas

Neste cenário, cinco filas são configuradas e duas filas não têm a razão de fila-buffers configurada.

Para determinar a alocação de buffer, a mesma lógica do Cenário 2 continua, mas você também deve dividir a Proporção Implícita Restante pelo número total de filas/filas implícitas que não têm proporção fila-buffers

Mapa de políticas em uso para este cenário:

<#root>

9500H(config)#

policy-map test4

9500H(config-pmap)#

class class1

9500H(config-pmap-c)#

priority level 1

9500H(config-pmap-c)#

queue-buffers ratio 20

9500H(config-pmap-c)#

class class2

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10 <-- no queue-buffers ratio statement for this class
```

```
9500H(config-pmap-c)#
```

```
class class3
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10 <-- no queue-buffers ratio statement for this class
```

```
9500H(config-pmap-c)#
```

```
class class4
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 10
```

```
9500H(config-pmap-c)#
```

```
class class-default
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 70
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 40
```

Resumo dos valores:

- Buffer base = 1200 - Como encontrar esse número é mostrado no Cenário 1
- Multiplicador de Usuário = 100% - Valor padrão

- Multiplicador Oculto = 4 - Padrão e não configurável pelo usuário
- Taxa de fila atual (class1) = 20% - Configurada no mapa de política com razão fila-buffers 20
- Taxa de fila atual (classe2) = Desconhecida, você deve resolver para esta - taxa de fila-buffers não está configurada para esta classe
- Taxa de fila atual (classe3) = Desconhecida, você deve resolver para esta - taxa de fila-buffers não está configurada para esta classe
- Proporção da fila atual (classe4) = 10% - Configurada no mapa de política com proporção de fila-buffers 10
- Taxa de fila atual (padrão de classe) = 40% - Configurada no mapa de política com taxa de fila-buffers 40

Calcular a taxa de buffer de fila que permanece:

Taxa implícita restante = (100 - Soma de taxas explicitamente configuradas).

Número de filas implícitas = 2 (classe2 e classe3 não têm relação de fila-buffers definida)

Soma das taxas configuradas = 20+40+10 = 70

Rácio implícito remanescente = 100 - 70 = 30

Alocação implícita de taxa de fila = [Taxa Implícita Restante] 30 / [Número de Filas Implícitas] 2 = 15

Calcular alocação de buffer de fila final:

Classe1 =

[Buffer Base] 1200 x [Taxa de Fila Atual (classe1)] (20/100) = 240 - fila de prioridade, nenhum cálculo adicional

Classe2 =

[Buffer Base] 1200 x [Alocação de taxa de fila implícita] (15/100) = 180 - Como a classe 2 não tem uma taxa de fila-buffers definida, o restante da taxa de fila-buffers de filas explícitas é compartilhado entre filas implícitas.

[Alocação de Buffer Base para esta fila] 180 x [Multiplicador oculto] 4 x [Multiplicador do Usuário] (100/100) = 720

Repita para as filas restantes:

Classe 3 = [

1200 x (15/100) = 180  
120 x 4 x (100/100) = 720  
]

Classe 4 = [

1200 x (10/100) = 120  
120 x 4 x (100/100) = 480  
]

Classe class-default = [

1200 x (40/100) = 480  
600 x 4 x (100/100) = 1920  
]

Resultado do mapa de políticas do teste 4 aplicado comparado com a previsão:

```
<#root>
```

```
9500H(config)#
```

```
interface tw1/0/3
```

```
9500H(config-if)#
```

```
service-policy output test4
```

```
9500H#
```

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12  
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
```

```

DrainFast:Disabled PortSoftStart:4 - 2880 BufferSharing:Disabled
DTS Hardmax Softmax PortSMin GblSMin PortStEnd QEnable
-----
0 1 5 240 8 240 0 0 0 0 6 3840 En
1 1 0 0 9

```

720

```

16 360 8 180 6 3840 En

```

<-- queue 1 and 2

were not configured with queue-buffers ratio

```

2 1 0 0 9

```

720

```

16 360 8 180 6 3840 En

```

<-- queue 1 and 2 get an equal share of leftover buffer ratio

```

3 1 0 0 11 480 16 240 8 120 6 3840 En
4 1 0 0 4 1920 16 960 8 480 6 3840 En
<snip>

```



Note: se o resultado da alocação de taxa de fila implícita não for um inteiro, uma parte igual não será possível. O resultado é arredondado para cima para filas anteriores no mapa de política e para baixo para filas posteriores. A soma final da taxa de buffers de fila alocada permanece 100, mas as filas implícitas nem sempre obtêm alocação igual devido ao requisito de resultado inteiro recém-descrito.

## Cenário 4 : A soma da razão de buffers de fila entre as filas é menor que 100

Neste cenário, cinco filas são configuradas, todas com proporção de fila-buffers. A soma total da razão de fila-buffers nas classes é menor que 100.

Nesse caso, a proporção de buffer não alocada é distribuída uniformemente entre as classes.

Semelhante ao cenário anterior, se o resultado dividido da relação fila-buffers restante não for um inteiro, a alocação final para cada fila será arredondada para cima ou para baixo e adicionada à

relação fila-buffers configurada.

Mapa de políticas em uso para este cenário:

```
<#root>
```

```
9500H(config)#
```

```
policy-map test5
```

```
9500H(config-pmap)#
```

```
class class1
```

```
9500H(config-pmap-c)#
```

```
priority level 1
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 10
```

```
9500H(config-pmap-c)#
```

```
class class2
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 10
```

```
9500H(config-pmap-c)#
```

```
class class3
```

```
9500H(config-pmap-c)#
```

```
bandwidth remaining percent 10
```

9500H(config-pmap-c)#

queue-buffers ratio 10

9500H(config-pmap-c)#

class class4

9500H(config-pmap-c)#

bandwidth remaining percent 10

9500H(config-pmap-c)#

queue-buffers ratio 10

9500H(config-pmap-c)#

class class-default

9500H(config-pmap-c)#

bandwidth remaining percent 70

9500H(config-pmap-c)#

queue-buffers ratio 12

Resumo dos valores:

- Buffer base = 1200 - Como encontrar esse número é mostrado no Cenário 1
- Multiplicador de Usuário = 100% - Valor padrão
- Multiplicador Oculto = 4 - Padrão e não configurável pelo usuário
- Taxa de fila atual (class1) = 10% - Configurada no mapa de política com relação fila-buffers 10
- Proporção da fila atual (classe2) = 10% - Configurada no mapa de política com proporção de fila-buffers 10
- Taxa de fila atual (classe3) = 10% - Configurada no mapa de política com a taxa de fila-buffers 10
- Proporção da fila atual (classe4) = 10% - Configurada no mapa de política com proporção de fila-buffers 10

- Taxa de fila atual (padrão de classe) = 12% - Configurado no mapa de política com razão fila-buffers 12
- Número total de filas = 5

Soma das Proporções Configuradas =  $10 + 10 + 10 + 10 + 12 = 52$

Taxa Restante do Buffer =  $100\% - 52\% = 48\%$

[Taxa de Buffer Restante] 48% / [Número Total de Filas] 5 = 9,6% adicionado por fila - Este não é um número inteiro, portanto sua aplicação final às filas deve ser arredondada para cima ou para baixo em uma base por fila

Para obter o número final da proporção de buffers de fila que o sistema usa, você deve adicionar 9 ou 10 à proporção de buffers de fila já configurada.

As classes superiores no mapa de políticas recebem o valor arredondado para cima, 10. As classes inferiores no mapa de políticas recebem o valor arredondado para baixo, 9.

Calcular alocação de buffer de fila final

Taxa Restante do Buffer = 48

Classe1 = [Buffer Base] x ([Taxa de fila atual (classe1) + Valor arredondado da taxa de buffer compartilhado restante])

Classe 1 =  $1200 \times ((10\% + 10\%)/100) = 240$  - fila de prioridade, nenhum cálculo adicional

Rácio de Buffer restante =  $(48 - 10) = 38$

Classe2 = [Buffer Base] x ([Taxa de fila atual (classe2) + Valor arredondado da taxa de buffer compartilhado restante])

Classe2 =  $1200 \times ((10\% + 10\%)/100) = 240$  - Continue a multiplicar isso pelos multiplicadores de usuário e sistema, pois essa não é uma fila de prioridade

Classe2 = [Alocação de Buffer Base para esta fila] 240 x [Multiplicador oculto] 4 x [Multiplicador do Usuário] (100/100) = 960 - resultado de softmax para esta fila

Rácio de buffer restante =  $(38 - 10) = 28$

Repita para as filas restantes:

Classe 3 = [

$$1200 \times ((10+10)/100) = 240$$

$$120 \times 4 \times (100/100) = 960$$

]

$$\text{Taxa de buffer restante} = (28 - 10) = 18$$

Classe 4 = [

$$1200 \times ((10+9)/100) = 240$$

$$120 \times 4 \times (100/100) = 912$$

]

$$\text{Taxa de buffer restante} = 9$$

Classe class-default= [

$$1200 \times ((12+9)/100) = 252$$

$$120 \times 4 \times (100/100) = 1008$$

]

$$\text{Taxa Restante do Buffer} = 0$$

Resultado do mapa de políticas test5 aplicado em comparação com a previsão:

<#root>

9500H#

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:5 - 1512 BufferSharing:Disabled
    DTS   Hardmax   Softmax   PortSMin   Glb1SMin   PortStEnd   QEnable
  -----
0    1    5    240    9
```

240

```
0 0 0 0 6 2016 En
1 1 0 0 10
```

960

```
16 480 8 240 6 2016 En
2 1 0 0 10
```

960

```
16 480 8 240 6 2016 En
3 1 0 0 11
```

912

```
16 456 8 228 6 2016 En
4 1 0 0 12
```

1008

```
16 504 8 252 6 2016 En
<snip>
```

## Cenário 5 : Determine o buffer flexível e rígido de uma classe de prioridade nível 2

Neste cenário, uma classe em um mapa de políticas é configurada com o nível de prioridade 2.

Ao contrário do nível de prioridade 1, no qual o softmax não é afetado por multiplicadores e é definido como hardmax, o nível de prioridade 2 permite que o softmax seja multiplicado, enquanto ele também tem uma alocação de buffer rígido (hardmax).

Mapa de políticas em uso para este cenário:

```
<#root>
```

```
9500H(config)#
```

```
policy-map test6
```

```
9500H(config-pmap)#
```

```
class class1
```

```
9500H(config-pmap-c)#
```

```
priority level 1
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 50 <-- 50 / 50 split between both queues
```

```
9500H(config-pmap-c)#
```

```
class class-default
```

```
9500H(config-pmap-c)#
```

```
priority level 2
```

```
<-- Priority level 2 in use now
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 50 <-- 50 / 50 split between both queues
```

Resultado do mapa de políticas de test6 aplicado:

```
<#root>
```

```
9500H#
```

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:5 - 3600 BufferSharing:Disabled
  DTS Hardmax Softmax PortSMin GblSMin PortStEnd QEnable
  -----
0 1 5 600 9
```

```
600
```

```
0 0 0 0 1 4800 En
```

```
<-- Softmax is equal to hardmax
```

```
1 1 5 600 10
```

```
2400
```

```
16 1200 0 0 1 4800 En
```

```
<-- Softmax is multiplied by Hidden Multiplier (400%) and User Multiplier (100% default)r
```

```
<snip>
```

Na saída mostrada anteriormente, o softmax da segunda fila tem 4 softmax da primeira fila. Isso ocorre porque o nível de prioridade 1 softmax especificamente não é afetado pelos multiplicadores softmax do sistema, mas o nível de prioridade 2 é.

Se você configurar um multiplicador softmax de usuário, somente a fila de nível de prioridade 2 será afetada:

```
<#root>
```

```
9500H(config)#
```

```
qos queue-softmax-multiplier 200
```

```
<#root>
```

```
9500H#
```

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:5 - 7200 BufferSharing:Disabled
  DTS Hardmax Softmax PortSMin GlbSMin PortStEnd QEnable
  -----
0 1 5 600 9 600 0 0 0 0 5 9600 En
```

```
<--- priority-level 1 queue unaffected by softmax multiplier
```

```
1 1 5 600 10
```

```
4800
```

```
8 1200 0 0 5 9600 En
```

```
<--- User multiplier increased to 200%, softmax for this queue doubles
```

```
<snip>
```

## Limite da fila

A configuração de limite de fila afeta a alocação final do buffer de fila

O mecanismo principal para influenciar a alocação de buffer de fila é a configuração `queue buffers-ratio` adicionada por fila em um mapa de políticas MQC.

No entanto, a alocação de buffer de fila é afetada por outras configurações.

O limite de fila define limites pelos quais você elimina uma classe de tráfego específica (via `Weighted Tail Drop`, WTD), que não é abordada neste documento.

Em uma circunstância específica, `queue-limit` modifica o Multiplicador oculto do sistema para filas de buffer soft - o que afeta a alocação geral de buffer soft para essa fila à qual `queue-limit` é aplicado.

Primeiro, entenda que o limite de fila pode ser configurado até 3 vezes por classe. Isso define até 3 limites para WTD por DSCP ou CoS.

Na próxima saída, apenas dois limites são definidos.

Aplique dois limites de fila a um mapa de políticas:

```
<#root>
```

```
9500H(config)#
```

```
policy-map test7
```

```
9500H(config-pmap)#
```

```
class class1
```

```
9500H(config-pmap-c)#
```

```
priority level 1
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 50
```

```
9500H(config-pmap-c)#
```

```
class class-default
```

```
9500H(config-pmap-c)#
```

```
priority level 2
```

```
9500H(config-pmap-c)#
```

```
queue-buffers ratio 50
```

```
9500H(config-pmap-c)#
```

```
queue-limit dscp af11 percent 10 <-- Tells system to drop af11 traffic at 10% queue utilization
```

```
9500H(config-pmap-c)#
```

```
queue-limit dscp af12 percent 50 <-- Tells system to drop af12 traffic at 50% queue utilization
```

Observe os resultados da alocação de buffer:

```
<#root>
```

```
9500H(config-pmap-c)#
```

```
interface tw1/0/3
```

```
9500H(config-if)#
```

```
service-policy output test7
```

9500H#

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:5 - 7200 BufferSharing:Disabled
  DTS Hardmax Softmax PortSMin GblSMin PortStEnd QEnable
  -----
0  1  5  600  9  600  0  0  0  0  5  9600  En
1  1  5  600 10
```

4800

```
8 1200 0 0 5 9600 En
```

<--- final result for queue that contains 2 queue-limit statements is 4800

<snip>

No próximo exemplo, uma terceira configuração queue-limit é adicionada à classe class-default.

Observe os resultados da alocação de buffer:

<#root>

```
9500H(config)#
```

```
policy-map test7
```

```
9500H(config-pmap)#
```

```
class class-default
```

```
9500H(config-pmap-c)#
```

```
queue-limit dscp af13 percent 100
```

9500H#

```
show platform hardware fed active qos queue config interface tw1/0/3
```

```
Asic:0 Core:1 DATA Port:22 GPN:103 LinkSpeed:0x12
```

```

AFD:Disabled FlatAFD:Disabled QoSMap:0 HW Queues: 176 - 183
  DrainFast:Disabled PortSoftStart:5 - 1800 BufferSharing:Disabled
  DTS  Hardmax  Softmax  PortSMin  G1b1SMin  PortStEnd  QEnable
  ----  -
0  1  5  600  9  600  0  0  0  0  5  2400  En
1  1  5  600  10

```

```
1200
```

```
32 1200 0 0 5 2400 En
```

```
<-- Softmax reduces by 400% from previous example
```

```
<snip>
```

Quando uma terceira configuração de limite de fila é adicionada a uma fila, o multiplicador de buffer flexível oculto do sistema de 400% é desabilitado para essa fila. No entanto, essa fila ainda respeita um qos queue-softmax-multiplier <percent> configurado pelo usuário.

## Informações Relacionadas

- [Plataformas de switching Cisco Catalyst 9000: White paper sobre QoS e enfileiramento](#)
- [Suporte técnico e downloads da Cisco](#)

## Sobre esta tradução

A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês ([link fornecido](#)) seja sempre consultado.