

# Analyse der Warteschlangenpufferzuweisung bei Catalyst Switches der Serie 9000

## Inhalt

---

### [Einleitung](#)

### [Voraussetzungen](#)

[Anforderungen](#)

[Verwendete Komponenten](#)

[Verwandte Produkte](#)

### [Hintergrundinformationen](#)

[Terminologie](#)

[Puffer](#)

[Algorithmus zur Zuweisung von Switch-Warteschlangenpuffern](#)

### [Konfigurieren](#)

[Szenario 1: Bestimmung des Pufferwerts für eine bestimmte Portgeschwindigkeit](#)

[Szenario 2: Bestimmen des weichen und harten Puffers für mehrere Warteschlangen und explizite Verhältnisse](#)

[Szenario 2.1: Vom Benutzer konfigurierter Multiplikator aus Szenario 2 ändern](#)

[Szenario 3: Ermitteln eines weichen und harten Puffers für mehrere Warteschlangen und eines impliziten Verhältnisses](#)

[Szenario 3.1: Bestimmen des weichen und harten Puffers für mehrere Warteschlangen und mehrere implizite Verhältnisse](#)

[Szenario 4: Die Summe der Warteschlangen-Puffer-Quotienten aus Warteschlangen ist kleiner als 100.](#)

[Szenario 5: Ermitteln des weichen und harten Puffers einer Klasse der Prioritätsstufe 2](#)

[Warteschlangenlimit](#)

### [Zugehörige Informationen](#)

---

## Einleitung

In diesem Dokument wird beschrieben, wie Sie die Pufferzuweisung für Datenverkehr in Warteschlangen auf Catalyst Switches der Serie 9000 vorhersagen können.

## Voraussetzungen

### Anforderungen

Cisco empfiehlt, dass Sie über Kenntnisse in folgenden Bereichen verfügen:

- Quality of Service (QoS)-Konzepte und -Prozesse für die Markierung, Warteschlangenverwaltung und Planung von Paketen
- Cisco MQC QoS-Konfiguration

## Verwendete Komponenten

Die Informationen in diesem Dokument basierend auf folgenden Software- und Hardware-Versionen:

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

Die Informationen in diesem Dokument beziehen sich auf Geräte in einer speziell eingerichteten Testumgebung. Alle Geräte, die in diesem Dokument benutzt wurden, begannen mit einer gelöschten (Nichterfüllungs) Konfiguration. Wenn Ihr Netzwerk in Betrieb ist, stellen Sie sicher, dass Sie die möglichen Auswirkungen aller Befehle kennen.

## Verwandte Produkte

Dieses Dokument kann auch mit folgenden Hardware- und Softwareversionen verwendet werden:

- Catalyst Switches der Serie 9200-9600
- Catalyst 9300X und 9400X
- Alle Cisco IOS XE Softwareversionen



Anmerkung: Das vorliegende Dokument gilt nicht für den 9500X oder 9600X, da diese eine andere ASIC- und QoS-Architektur verwenden.

---

## Hintergrundinformationen

Eine technische Übersicht über die QoS auf Catalyst Switches der Serie 9000 finden Sie unter: [Whitepaper: Catalyst 9000 QoS und Warteschlangen](#)

Häufig muss die Pufferzuweisung als Reaktion auf unerwünschte Ausgabeverluste für eine bestimmte Datenverkehrs-kategorie angepasst werden. Weitere Informationen zum Diagnostizieren

und Beheben von Ausgabefehlern bei Catalyst Switches der Serie 9000 finden Sie in diesem Artikel: [Fehlerbehebung für verworfene Ausgangspakete auf Switches der Catalyst 9000-Serie](#)

## Terminologie

QoS	Quality of Service	Ein Konzept/eine Gruppe verwandter Funktionen, die sich auf die Klassifizierung, Markierung, Warteschlangeneinrichtung und Planung des ein- und ausgehenden Datenverkehrs eines Netzwerkgeräts beziehen
DSCP	Differentiated Services Code Point	Ein Verkehrsklassifizierungsmechanismus, der im IP-Header eines Pakets enthalten ist
CoS	Class of Service	Ein Verkehrsklassifizierungsmechanismus, der im Ethernet-Frame-Header eines Pakets enthalten ist.
RAUM	Zugriffskontrolleintrag	Eine einzelne Regel oder Zeile in einer Zugriffskontrollliste (ACL)
ACL	Zugriffskontrollliste	Eine Gruppe von Zugriffskontrolleinträgen (Access Control Entries, ACEs), die von verschiedenen Funktionen verwendet werden, um Datenverkehr abzugleichen und eine Aktion auszuführen.
ASIC	Anwendungsspezifische integrierte Schaltung	Ein Computerchip, der entwickelt wurde, um eine bestimmte Aufgabe oder eine Reihe von Aufgaben mit hoher Effizienz auszuführen.
UADP	Unified Access-Datenebene	Der in Catalyst Switches der Serie 9000 verwendete Cisco ASIC führt viele Netzwerkpaketverarbeitungsaufgaben aus.
PBC	Paket-Puffer-Komplex	Cisco UADP ASIC-Subsystem, das als zentraler Paketpuffer für die Verarbeitung, Warteschlangeneinrichtung und Planung von Paketen dient
AQM	Aktive Warteschlangenverwaltung	Cisco UADP ASIC-Subsystem, das Datenverkehrswarteschlangen verwaltet und Aktionen für Netzwerkports plant.

DTS	Dynamische Grenzwerte und Skalierbarkeit	Cisco UADP ASIC-Technologie zur dynamischen Anpassung und Skalierung von Puffern über die Ports zur Optimierung der Hardwarenutzung
-----	--	---

## Puffer

Puffer sind Speicher, die dazu dienen, vorübergehende Datenspitzen zu absorbieren, wenn die an einen Port gesendeten oder weitergeleiteten Daten die Fähigkeit der Ports übersteigen, Daten über das Kabel zu übertragen. Ein Port hat eine feste Übertragungsrate und entfernt Daten aus der Warteschlange. Im Prinzip ist ein Puffer einfach ein Ort, an dem Daten gespeichert oder in eine Warteschlange gestellt werden, bis sie über die Schnittstelle übertragen werden.

Auf Catalyst Switches der Serie 9000 hat das Wort "Puffer" zwei Verwendungen. Der Systempuffer als Ganzes wird auch als Paketpufferkomplex (PBC) des ASIC bezeichnet. Die Wortpuffer können auch kleine Einheiten des PBC referenzieren. Den Ports wird auf Warteschlangensbasis ein Puffer zugewiesen. Mit anderen Worten wird einer Port-Warteschlange eine Menge kleiner einzelner Puffer aus dem Gesamtsystempuffer zugewiesen.

Auf ASIC-basierten Cisco UADP-Plattformen enthält ein Puffer bis zu 256 Byte an Daten, und Puffer sind miteinander verknüpft, um Frames mit mehr als 256 Byte darzustellen.

## Algorithmus zur Zuweisung von Switch-Warteschlangenspuffern

- Einem Port ist ein hartcodierter Pufferwert zugewiesen, der vom Benutzer oder System skaliert werden kann (die Basisnummer kann jedoch nicht geändert werden, was wichtig ist, um die Ergebnisse vorherzusagen).
- Der Pufferwert für einen bestimmten Port-Typ hängt von der spezifischen Kombination aus Hardware (Switch-Modell), Softwareversion und Portgeschwindigkeit ab. Ein Gigabit-Ethernet-Port auf dem Catalyst C9500-25Y4C beginnt beispielsweise mit einem anderen Puffergrundwert als ein Gigabit-Ethernet-Port auf dem Catalyst 9200L.
- Dieser Basiswert wird je nach Konfiguration auf N Warteschlangen aufgeteilt.

Die endgültige Berechnung des verfügbaren Puffers pro Warteschlange wird durch folgende Faktoren beeinflusst:

- Anzahl der konfigurierten Warteschlangen
- Das Warteschlangen-Puffer-Verhältnis. Mit diesem Befehl können Sie einstellen, wie der

Basispuffer auf die konfigurierten Port-Warteschlangen aufgeteilt wird.

- Ein vom Benutzer konfigurierter Puffermultiplikator (konfiguriert als QoS-Warteschlangen-Softmax-Multiplikator <Prozent>)
- Ein vom System definierter Puffermultiplikator (nicht vom Benutzer konfigurierbar, 4-fach oder 400 %) - während dies standardmäßig angewendet wird, gibt es einige Umstände, unter denen seine Entfernung so wichtig ist, um diesen 4x-Wert zu verstehen, existiert, um den Zuweisungsalgorithmus vorherzusagen
- Beide Multiplikatoren wirken sich nur auf sog. Soft-Puffer aus

Weiche Puffer sind Puffer, die über mehrere Ports verteilt werden. Diese Puffer werden als "weich" bezeichnet, da sie für den Port nicht garantiert sind.

Das System ordnet weiche Puffer absichtlich zu. Auf diese Weise kann ein Port bei Bedarf eine große Anzahl von Puffern verwenden. Da jedoch mehr Ports Puffer benötigen, werden alle Ports und Warteschlangen im Rahmen des Cisco UADP ASIC DTS-Prozesses dynamisch und angemessen verkleinert.

Zusammenfassend lässt sich sagen, dass weiche Puffer - in den Ausgaben als softmax bezeichnet - ein opportunistischer Maximalwert sind. Ein Port verwendet nur dann den vollen softmax-Wert, wenn die entsprechende Puffermenge aus dem gesamten Systempuffer verfügbar ist. Wenn die Pufferanforderung an andere Ports und Warteschlangen ansteigt, ist der für den Port maximal verfügbare Puffer geringer.

Harte Puffer sind Puffer, die explizit für einen Port reserviert sind und nicht vom DTS-Prozess beeinflusst werden. Da feste Puffer garantierte Puffer sind, übersteigt die Summe der den Ports zugewiesenen harten Puffer niemals das für diese harten Puffer reservierte PBC-Segment.

Die Mechanismen, die die aktive Skalierung von Soft-Buffer steuern, werden als DTS (Dynamic Threshold and Scale) bezeichnet und im [Whitepaper](#) zu [QoS](#) der [Catalyst Serie 9000](#) beschrieben.

Die Größe der PBC-Segmente, die für die Warteschlangen-Hard- und Soft-Puffer reserviert sind, ändert sich dynamisch, wenn Sie das System konfigurieren. Sie können in dieser Ausgabe als AQM GlobalSoftLimit und GlobalHardLimit angezeigt werden:

```
<#root>
```

```
C9500#
```

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

```
-----  
AQM Global counters
```

GlobalHardLimit:

18072

| GlobalHardBufCount: 0  
GlobalSoftLimit:

37224

| GlobalSoftBufCount: 0

C9500#

show platform hardware fed active qos queue config interface tw1/0/1

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	GlbSMin	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	GlbISMin	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



Anmerkung: Beachten Sie die Zunahme von GlobalHardLimit und die proportionale Abnahme von GlobalSoftLimit.

---

Wenn Sie außerdem die Prioritätsstufe 1 konfigurieren, ist die Softmax für diese Warteschlange statisch genau gleich der hardmax. Sie können den festen Puffer nur für eine Warteschlange der Prioritätsstufe 1 ändern.

Die Änderung in GlobalHardLimit und GlobalSoftLimit ist gleich 720. Dies ist auch gleich der Änderung in hardmax nach der Konfiguration.

In den Szenarien in diesem Dokument wird erläutert, wie Sie die Softmax- und Hardmax-Zuweisungen für mehrere Richtlinienzuordnungskonfigurationen berechnen und vorhersagen.

## Konfigurieren

### Szenario 1: Bestimmung des Pufferwerts für eine bestimmte Portgeschwindigkeit

Ein Endpufferwert einer Warteschlange ist zum Teil eine Funktion eines Basiswerts, der zunächst der Warteschlange zugeordnet wird. Dieser wird dann später bei weichen Puffern multipliziert.

Die Multiplikationsfaktoren in Kombination mit anderen impliziten Verhaltensweisen machen die Bestimmung eines Endwerts für eine bestimmte Warteschlange mit einer bestimmten Konfiguration zu einer Herausforderung.

Der erste Schritt zur Klärung der resultierenden Warteschlangenpufferzuordnung besteht in der Bestimmung des Basispufferwerts.

Dazu nutzen Sie eine Prioritätswarteschlange, die einen festen Puffer erhält, der entweder proportional zur Anzahl der Warteschlangen oder zum konfigurierten Verhältnis von Warteschlangen und Puffern ist.

Mit einer bestimmten Konfiguration können Sie explizit die Menge an Basispuffer ableiten, die einer bestimmten Portgeschwindigkeit zugewiesen ist.

Konfiguration und Zuweisung des gesamten Puffers zu einer einzelnen, nicht multiplizierten Warteschlange (Prioritätsstufe 1-Warteschlange)

In diesem Beispiel wird die class-default-Klasse verwendet, um den gesamten Datenverkehr abzugleichen, da keine anderen Klassen konfiguriert sind.

```
<#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
```

Die Konfiguration im vorherigen Beispiel führt die folgenden Aktionen aus:

- Die Prioritätsstufe 1 weist das System an, einen strikten Zeitplan für diese Warteschlange/Klasse auszuführen und dieser Warteschlange/Klasse einen festen Puffer zuzuweisen.
  - Die Planung von Paketen geht über den Rahmen dieses Dokuments hinaus. Eine strikte Warteschlange wird jedoch in der Regel vor allen anderen Warteschlangen verwaltet.
- Das Warteschlangen-Puffer-Verhältnis 100 weist dieser Warteschlange/Klasse 100/100 oder 100 % des verfügbaren Basispuffers zu.

In einer Richtlinienzuordnung mit mehr als einer Klasse können Sie einer einzelnen Klasse nicht 100 % des Puffers zuweisen. Sie müssen mindestens 1/100 oder 1 % einer Klasse zuweisen.

In einer Richtlinie mit nur einer Klasse gibt es nur eine Klasse, und Sie können ihr den gesamten Puffer zuweisen.

Wie bereits erwähnt, erhält eine Prioritätswarteschlange feste Puffer, die der Verteilung des Basispuffers gemäß dem konfigurierten Verhältnis von Warteschlangen zu Puffern entsprechen. Ein harter Puffer unterliegt keinem Multiplikator.

Hard Buffer wird in Ausgaben unter einer Spalte mit dem Titel Hardmax beobachtet.

Jetzt haben Sie eine einzige Datenverkehrsklasse mit Puffern, die keinem Multiplikator unterliegen. Damit können Sie explizit eine Pufferzuweisung für diese Portgeschwindigkeit ableiten (und nur diese Portgeschwindigkeit auf dieser Plattform, andere unterscheiden sich), da Basispuffer und hardmax gleich sind.

Basispuffer = ?

Warteschlangenverhältnis 1 = 100/100 = 1

Hardmax für diese Warteschlange = Basispuffer x Warteschlangenverhältnis 1

$X = Y \times 1$

$X/1 = Y$

$X = Y$

X = Y = Hardmax = Basispuffer = 1200 (siehe Beispiel 2).

In diesem Beispiel wird "policy-map test1" als Ausgabeservice-Richtlinie auf eine Schnittstelle angewendet.

<#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	Glb1SMin	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>

Wie gezeigt, beträgt die maximale Hardmax für diese Prioritätswarteschlange mit 100 %

zugewiesenem Puffer 1200.

Da `hardmax` ein vollständig nicht multiplizierter/nicht skaliertes Wert ist und 100 % des Puffers für diese Warteschlange konfiguriert sind, beträgt die Puffergrundzuweisung für dieses Switch-Modell, die Softwareversion und die spezifische Portgeschwindigkeit 1200.

Die anderen Port-Geschwindigkeiten auf diesem Switch und andere Switches für die gleiche Port-Geschwindigkeit erhalten unterschiedliche Pufferreserven. Diese Basiszuweisung ist vom Benutzer nicht konfigurierbar und muss durch Beobachtung abgeleitet werden.

Weitere Szenarien in diesem Dokument beginnen mit demselben Switch, derselben Software und derselben Portgeschwindigkeit. Sie alle nehmen also eine Basiszuweisung von 1200 an, um die endgültige Pufferzuweisung zu berechnen.



Anmerkung: Beachten Sie, dass der `softmax`-Wert im vorherigen Beispiel ebenfalls 1200 beträgt.

---

Eine Prioritätsstufe-1-Warteschlange hat `softmax` genau gleich ihrem `hardmax`. Dies ist beabsichtigt und kann nicht vom Benutzer konfiguriert werden.

Außerdem wird dieser spezielle Fall der `softmax`-Zuordnung durch die nachfolgend dargestellten `softmax`-Multiplikatoren nicht beeinflusst. Nur eine Warteschlange der Prioritätsstufe 1 weist dieses Verhalten für `softmax` auf, was beabsichtigt ist.

## Szenario 2: Bestimmen des weichen und harten Puffers für mehrere Warteschlangen und explizite Verhältnisse

In diesem Szenario wird eine zusätzliche Warteschlange hinzugefügt. Diese Warteschlange verwendet nicht die Prioritätsstufe 1, sodass `softmax` mit Multiplikatoren skaliert werden kann.

Ein Multiplikator ist vom Benutzer konfiguriert und der andere ein versteckter/nicht konfigurierbarer Multiplikator.

Kombinieren Sie diese Multiplikatoren mit dem für diesen Port abgeleiteten Basispuffer, in diesem Fall 1200 gemäß Szenario 1.

Algorithmisch:

Aktuelles Warteschlangenverhältnis = Warteschlangenpufferverhältnis für die vorherzusagende Warteschlange/Klasse

Versteckter Multiplikator = 400 %

User Multiplier = Prozentwert, den Sie in qos queue-softmax-multiplier <percentage> konfigurieren. Der Standardwert ist 100 %.

Softmax = ( Basispuffer x (aktuelles Warteschlangenverhältnis / 100)) x versteckter Multiplikator x (Benutzermultiplikator / 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
```

Zusammenfassung der Werte:

- Basispuffer = 1200 - Wie diese Zahl zu finden ist, wird in Szenario 1 gezeigt
- Benutzer-Multiplikator = 100% - Standardwert, dieser wurde noch nicht geändert
- Versteckter Multiplikator = 4 - Standard und nicht vom Benutzer konfigurierbar
- Aktuelles Warteschlangenverhältnis (class1) = 50 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 50
- Aktuelles Warteschlangenverhältnis (Klassenstandard) = 50 % - in der Richtlinienzuordnung mit Warteschlangenerverhältnis 50 konfiguriert

Bestimmen der Klasse-1-Pufferzuweisung:

Da class1 eine Prioritätswarteschlange ist, empfängt sie hardmax (hard buffer) und einen Sonderfall von softmax, der nicht von Multiplikatoren beeinflusst wird.

Class1 hardmax = (Basispuffer x aktuelles Warteschlangenverhältnis (class1) / 100)

Class1 hardmax = 1200 x (50/100) = 600 - aufgrund eines Sonderfalls einer

Prioritätswarteschlange alle Mathematik anhalten, Ergebnis hardmax zuordnen. Softmax ist in der Regel Hardmax für die Prioritätsstufe 1.

Bestimmen Sie die standardmäßige Pufferzuweisung für die Klasse:

Class class-default = (Basispuffer x (Aktuelles Warteschlangenverhältnis (class-default / 100)) x

Versteckter Multiplikator x (Benutzermultiplikator / 100)

Klasse class-default = [

[Basispuffer] 1200 x [Aktuelles Warteschlangenverhältnis] (50/100) = 600

[vorheriges Ergebnis] 600 x [Versteckter Multiplikator] 4 x [User Multiplier] (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 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 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>

Endergebnis Q0 - Hardmax: 600 Softmax: 600. Q1 - Softmax: 2400

Szenario 2.1: Vom Benutzer konfigurierter Multiplikator aus Szenario 2 ändern

Dieses Szenario beginnt wie Szenario 2, mit der Ausnahme, dass Sie jetzt qos queue-softmax-multiplier 1200 konfigurieren.

Dadurch werden die softmax-Puffer in der aktuellen Konfiguration mit 1200 % bzw. einem Faktor von 12 multipliziert.

Zusammenfassung der Werte:

- Basispuffer = 1200 - Wie diese Zahl zu finden ist, wird in Szenario 1 gezeigt
- User Multiplier = 1200% - Dies ist der in diesem Szenario geänderte Wert über qos queue-softmax-multiplier 1200
- Versteckter Multiplikator = 4 - Standard und nicht vom Benutzer konfigurierbar
- Aktuelles Warteschlangenverhältnis (class1) = 50 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 50
- Aktuelles Warteschlangenverhältnis(class-default) = 50 % - konfiguriert in der Richtlinienzuordnung mit Warteschlangen-Puffer-Verhältnis 50

Bestimmen der Klasse-1-Pufferzuweisung:

Da class1 eine Prioritätswarteschlange ist, empfängt sie hardmax (hard buffer) und einen Sonderfall von softmax, der nicht von Multiplikatoren beeinflusst wird.

Class1 hardmax = (Basispuffer x aktuelles Warteschlangenverhältnis (class1) / 100)

Class1 hardmax =  $1200 \times (50/100) = 600$  - aufgrund eines Sonderfalls einer

Prioritätswarteschlange alle Mathematik anhalten, Ergebnis hardmax zuordnen. Softmax ist in der Regel Hardmax für die Prioritätsstufe 1.

Bestimmen Sie die standardmäßige Pufferzuweisung für die Klasse:

Class class-default = (Basispuffer x (Aktuelles Warteschlangenverhältnis (class-default / 100)) x Versteckter Multiplikator x (Benutzermultiplikator / 100)

Klasse class-default =[

[Basispuffer] 1200 x [Aktuelles Warteschlangenverhältnis] (50/100) = 600

[vorheriges Ergebnis] 600 x [Verborgener Multiplikator] 4 x [Benutzer-Multiplikator] (1200/100) = 28800

]

Konfigurieren Sie qos queue-softmax-multiplier 1200 und beobachten Sie Änderungen an softmax (softmax ist ein maximaler Pufferwert für diese Warteschlange, dynamisch skaliert basierend auf der aktuellen Gesamtpuffernutzung):

<#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>
```

Endergebnis Q1 - Hardmax: 600, Softmax: 600. Q2 - Softmax: 28800

### Szenario 3: Ermitteln eines weichen und harten Puffers für mehrere Warteschlangen und eines impliziten Verhältnisses

In diesem Szenario werden fünf Warteschlangen konfiguriert, aber nur vier weisen ein explizit definiertes Verhältnis zwischen Warteschlangen und Puffern auf. Der diesen Warteschlangen zugewiesene Puffer entspricht dem in den vorherigen Beispielen.

Die nicht konfigurierte Warteschlange erhält die Differenz zwischen der Summe aller konfigurierten Warteschlangenkoeffizienten und 100.

Summe der explizit konfigurierten Verhältnisse = (Q0-Pufferverhältnis) + (Q1-Pufferverhältnis) ... (letztes Pufferverhältnis) - bis zu 8 Warteschlangen werden auf Catalyst Switches der Serie 9000 unterstützt, sodass Sie bis zu 8 Verhältnisse hinzufügen können

Überbleibsel des impliziten Verhältnisses = (100 - Summe der explizit konfigurierten Verhältnisse)

Implicit Ratio Left ist der Wert, der einer Warteschlange zugewiesen wird, für die kein Warteschlangen-Puffer-Verhältnis konfiguriert wurde.

Die Richtlinienzuweisung wird für dieses Szenario verwendet:

```
<#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
```

Zusammenfassung der Werte:

- Basispuffer = 1200 - Wie diese Zahl zu finden ist, wird in Szenario 1 gezeigt
- Benutzer-Multiplikator = 100% - Standardwert, dieser wurde noch nicht geändert
- Versteckter Multiplikator = 4 - Standard und nicht vom Benutzer konfigurierbar
- Aktuelles Warteschlangenverhältnis (class1) = 20 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 20
- Aktuelles Warteschlangenverhältnis (class2) = Unbekannt, Sie müssen dieses Problem lösen - Warteschlangenspufferverhältnis ist für diese Klasse nicht konfiguriert
- Aktuelles Warteschlangenverhältnis (class3) = 10 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 10
- Aktuelles Warteschlangenverhältnis (class4) = 10 % - In der Richtlinienzuordnung mit Warteschlangenspufferverhältnis 10 konfiguriert
- Aktuelles Warteschlangenverhältnis(class-default) = 40 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 40

Das verbleibende Warteschlangenspufferverhältnis berechnen:

Überbleibsel des impliziten Verhältnisses = (100 - Summe der explizit konfigurierten Verhältnisse)

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

Aktuelles Warteschlangenverhältnis (class2) = 20

Endgültige Warteschlangenpufferzuweisung berechnen

Class1 = [Base Buffer] 1200 x [Current Queue Ratio(class1)] (20/100) = 240 - Priority Queue, keine weitere Berechnung

Klasse 2 = [

[Basispuffer] 1200 x [Aktuelles Warteschlangenverhältnis (Klasse 2)] (10/100) = 240 -

Basispufferzuweisung für diese Warteschlange, muss aber multipliziert werden, um softmax für eine Warteschlange ohne Priorität zu erhalten.

[Basis-Puffer-Zuweisung für diese Warteschlange] 120 x [Ausgeblendeter Multiplikator] 4 x [Benutzer-Multiplikator] (100/100) = 960

]

Wiederholen Sie den Vorgang für verbleibende Warteschlangen:

Klasse 3 = [

1200 x (10/100) = 120

120 x 4 x 100/100 = 480

]

Klasse 4 = [

1200 x (10/100) = 120

120 x 4 x 100/100 = 480

]

Klasse class-default = [

1200 x (40/100) = 480

600 x 4 x (100/100) = 1.920

]

Ergebnis der angewendeten test3-Richtlinienzuweisung im Vergleich zur Prognose:

<#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 GblSMin 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>
```

Endergebnis Q0 - Hardmax: 240, Softmax: 240. Q1 - Softmax: 960, 2. Quartal - Softmax: 480, 3. Quartal - Softmax: 480, 4. Quartal - Softmax: 480

Szenario 3.1: Bestimmen des weichen und harten Puffers für mehrere Warteschlangen und mehrere implizite Verhältnisse

In diesem Szenario werden fünf Warteschlangen konfiguriert, und für zwei Warteschlangen ist das Verhältnis von Warteschlangen zu Puffern nicht konfiguriert.

Zur Bestimmung der Pufferzuweisung wird die gleiche Logik aus Szenario 2 fortgesetzt. Sie müssen jedoch auch den Überbleibsel des impliziten Verhältnisses durch die Gesamtzahl impliziter Warteschlangen/Warteschlangen dividieren, die kein Verhältnis zwischen Warteschlangen und Puffern aufweisen.

Die Richtlinienzuweisung wird für dieses Szenario verwendet:

```
<#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
```

Zusammenfassung der Werte:

- Basispuffer = 1200 - Wie diese Zahl zu finden ist, wird in Szenario 1 gezeigt
- Benutzermultiplikator = 100 % - Standardwert
- Versteckter Multiplikator = 4 - Standard und nicht vom Benutzer konfigurierbar
- Aktuelles Warteschlangenverhältnis (class1) = 20 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 20
- Aktuelles Warteschlangenverhältnis (class2) = Unbekannt, Sie müssen dieses Problem lösen - Warteschlangenpufferverhältnis ist für diese Klasse nicht konfiguriert
- Current Queue Ratio(class3) = Unbekannt, Sie müssen dieses Problem lösen - Queue-Puffer-Verhältnis ist für diese Klasse nicht konfiguriert
- Aktuelles Warteschlangenverhältnis (class4) = 10 % - In der Richtlinienzuordnung mit Warteschlangenpufferverhältnis 10 konfiguriert
- Aktuelles Warteschlangenverhältnis (Klassenstandard) = 40 % - In der Richtlinienzuordnung mit Warteschlangenpufferverhältnis 40 konfiguriert

Das verbleibende Warteschlangenpufferverhältnis berechnen:

Überbleibsel des impliziten Verhältnisses =  $(100 - \text{Summe der explizit konfigurierten Verhältnisse})$

Anzahl der impliziten Warteschlangen = 2 (class2 und class3 haben kein definiertes Verhältnis zwischen Warteschlangen und Puffern)

Summe der konfigurierten Kennzahlen =  $20+40+10 = 7$

Ungültiger Quotient =  $100 - 70 = 30$

Implizite Zuweisung des Warteschlangenverhältnisses = [Überbleibsel des impliziten Verhältnisses]  $30 /$  [Anzahl impliziter Warteschlangen]  $2 = 15$

Berechnen der endgültigen Warteschlangenpufferzuordnung:

Klasse 1 =

[Basispuffer]  $1200 \times$  [Aktuelles Warteschlangenverhältnis (Klasse 1)]  $(20/100) = 240$  -  
Prioritätswarteschlange, keine weitere Berechnung

Klasse 2 =

[Basispuffer]  $1200 \times$  [Zuweisung des impliziten Warteschlangenverhältnisses]  $(15/100) = 180$  - Da Klasse 2 kein definiertes Verhältnis zwischen Warteschlangen und Puffern aufweist, wird das Verhältnis zwischen Warteschlangen und Puffern aus expliziten Warteschlangen auf implizite Warteschlangen aufgeteilt.

[Basis-Puffer-Zuweisung für diese Warteschlange]  $180 \times$  [Verborgener Multiplikator]  $4 \times$  [Benutzer-Multiplikator]  $(100/100) = 720$

Wiederholen Sie den Vorgang für verbleibende Warteschlangen:

Klasse 3 = [

$1200 \times (15/100) = 180$   
 $120 \times 4 \times (100/100) = 720$

]

Klasse 4 = [

$1200 \times (10/100) = 120$   
 $120 \times 4 \times 100/100 = 480$

]

Klasse class-default = [

1200 x (40/100) = 480

600 x 4 x (100/100) = 1.920

]

Ergebnis der angewendeten test4-Richtlinienzuordnung im Vergleich zur Vorhersage:

<#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 Glb1SMin PortStEnd QEnable

	DTS	Hardmax	Softmax	PortSMin	Glb1SMin	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>
```



Anmerkung: Wenn das Ergebnis der impliziten Zuweisung des Warteschlangenverhältnisses keine ganze Zahl ist, ist eine gleichmäßige Verteilung nicht möglich. Das Ergebnis wird für Warteschlangen früher in der Richtlinienzuordnung und für spätere Warteschlangen aufgerundet. Die endgültige Summe der zugewiesenen Warteschlangenpufferquote bleibt 100, aber implizite Warteschlangen erhalten aufgrund der eben beschriebenen Anforderungen an das Integerergebnis nicht immer die gleiche Zuweisung.

#### Szenario 4: Die Summe der Warteschlangen-Puffer-Quotienten aus Warteschlangen ist kleiner als 100.

In diesem Szenario werden fünf Warteschlangen konfiguriert, alle mit einem Verhältnis von Warteschlangen-Puffern. Die Summe der Warteschlangen-Puffer-Quotienten zwischen den Klassen beträgt weniger als 100.

In diesem Fall wird das nicht zugewiesene Pufferverhältnis gleichmäßig über die Klassen verteilt.

Ähnlich wie im vorherigen Szenario wird die endgültige Zuordnung zu jeder Warteschlange aufgerundet oder abgerundet und zum von Ihnen konfigurierten Verhältnis der Warteschlangenpuffer hinzugefügt, wenn das geteilte Ergebnis des Verhältnisses zwischen verbleibenden Warteschlangen und Puffern keine ganze Zahl ist.

Die Richtlinienzuweisung wird für dieses Szenario verwendet:

```
<#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

### Zusammenfassung der Werte:

- Basispuffer = 1200 - Wie diese Zahl zu finden ist, wird in Szenario 1 gezeigt
- Benutzermultiplikator = 100 % - Standardwert
- Versteckter Multiplikator = 4 - Standard und nicht vom Benutzer konfigurierbar
- Aktuelles Warteschlangenverhältnis (class1) = 10 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 10
- Aktuelles Warteschlangenverhältnis (class2) = 10 % - In der Richtlinienzuordnung mit Warteschlangenpufferverhältnis 10 konfiguriert
- Aktuelles Warteschlangenverhältnis (class3) = 10 % - in der Richtlinienzuordnung mit Warteschlangenpufferverhältnis 10 konfiguriert
- Aktuelles Warteschlangenverhältnis (class4) = 10 % - In der Richtlinienzuordnung mit Warteschlangenpufferverhältnis 10 konfiguriert
- Aktuelles Warteschlangenverhältnis (Klassenstandard) = 12 % - In der Richtlinienzuordnung konfiguriert mit Warteschlangen-Puffer-Verhältnis 12
- Gesamtzahl der Warteschlangen = 5

Summe der konfigurierten Übersetzungen =  $10 + 10 + 10 + 10 + 12 = 52$

Restpufferquote =  $100 \% - 52 \% = 48 \%$

[Verbleibendes Pufferverhältnis]  $48 \% /$  [Gesamtzahl der Warteschlangen]  $5 = 9,6 \%$  pro Warteschlange hinzugefügt - Dies ist keine ganze Zahl, daher muss die endgültige Anwendung auf die Warteschlangen auf Warteschlangensbasis aufgerundet oder heruntergerundet werden.

Um das endgültige Verhältnis von Warteschlangenpuffern zu erhalten, die vom System verwendet werden, müssen Sie dem bereits konfigurierten Verhältnis von Warteschlangenpuffern 9 oder 10

hinzufügen.

Höhere Klassen in der Richtlinienzuordnung erhalten den aufgerundeten Wert, 10. Niedrigere Klassen in der Richtlinienzuordnung erhalten den aufgerundeten Wert, 9.

Endgültige Warteschlangenpufferzuweisung berechnen

Verbleibendes Pufferverhältnis = 48

Klasse1 = [Basispuffer] x ([Aktuelles Warteschlangenverhältnis (Klasse1) + Aufgerundeter Wert des verbleibenden gemeinsamen Pufferverhältnisses])

Klasse1 =  $1200 \times ((10 \% + 10 \%)/100) = 240$  - Prioritätswarteschlange, keine weitere Berechnung

Pufferverhältnis übrig =  $(48 - 10) = 38$

Klasse2 = [Basispuffer] x ([Aktuelles Warteschlangenverhältnis (Klasse2) + Aufgerundeter Wert des verbleibenden gemeinsamen Pufferverhältnisses])

Klasse2 =  $1200 \times ((10 \% + 10 \%)/100) = 240$  - Multiplizieren Sie dies mit Benutzer- und Systemmultiplikatoren, da es sich nicht um eine Prioritätswarteschlange handelt.

Klasse2 = [Basis-Puffer-Zuweisung für diese Warteschlange]  $240 \times$  [Ausgeblendeter Multiplikator]  $4 \times$  [Benutzer-Multiplikator]  $(100/100) = 960$  - Softmax-Ergebnis für diese Warteschlange

Restliches Pufferverhältnis =  $(38 - 10) = 28$

Wiederholen Sie den Vorgang für verbleibende Warteschlangen:

Klasse 3 = [

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

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

]

Restliches Pufferverhältnis =  $(28 - 10) = 18$

Klasse 4 = [

1200 x (10+9)/100 = 240  
120 x 4 x (100/100) = 912  
]

Pufferverhältnis übrig = 9

Class class-default= [

1200 x (12+9)/100 = 252  
120 x 4 x (100/100) = 1.008  
]

Pufferverhältnis übrig = 0

Ergebnis der angewendeten test5-Richtlinienzuweisung im Vergleich zur Prognose:

<#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 GlbSMin 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>
```

## Szenario 5: Ermitteln des weichen und harten Puffers einer Klasse der Prioritätsstufe 2

In diesem Szenario wird eine Klasse in einer Richtlinienzuordnung mit der Prioritätsstufe 2 konfiguriert.

Im Gegensatz zur Prioritätsstufe 1, bei der softmax nicht durch Multiplikatoren beeinflusst wird und gleich hardmax gesetzt ist, erlaubt die Prioritätsstufe 2 das Multiplizieren von softmax, während sie auch eine harte Pufferzuordnung (hardmax) aufweist.

Die Richtlinienzuweisung wird für dieses Szenario verwendet:

```
<#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
```

Ergebnis der angewendeten Test6-Richtlinienzuweisung:

```
<#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 Glb1SMin 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>
```

In der zuvor gezeigten Ausgabe hat die zweite Warteschlange softmax 4 so viel softmax wie die erste Warteschlange. Dies liegt daran, dass die Prioritätsstufe 1 softmax von den System-softmax-Multiplikatoren ausdrücklich nicht betroffen ist, die Prioritätsstufe 2 jedoch.

Wenn Sie einen Benutzer-Softmax-Multiplikator konfigurieren, ist nur die Warteschlange der Prioritätsstufe 2 betroffen:

```
<#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>
```

## Warteschlangenlimit

Die Warteschlangengrenzwertkonfiguration wirkt sich auf die endgültige

Warteschlangenpufferzuweisung aus.

Der primäre Mechanismus zur Beeinflussung der Warteschlangenpufferzuweisung ist die Konfiguration des Warteschlangenpufferverhältnisses, die pro Warteschlange in einer MQC-Richtlinienzuordnung hinzugefügt wird.

Die Zuweisung von Warteschlangenpuffern wird jedoch durch andere Konfigurationen beeinflusst.

Die Warteschlangenbeschränkung definiert Schwellenwerte, durch die eine bestimmte Datenverkehrs-kategorie (Weighted Tail Drop, WTD) verworfen wird, die in diesem Dokument nicht behandelt wird.

Unter bestimmten Umständen ändert `queue-limit` den System-Hidden Multiplier für Soft-Buffer-Warteschlangen, was sich auf die gesamte Soft-Buffer-Zuweisung für die Warteschlange auswirkt, auf die `queue-limit` angewendet wird.

Beachten Sie zunächst, dass die Warteschlangenbeschränkung bis zu dreimal pro Klasse konfiguriert werden kann. Hiermit werden bis zu drei Grenzwerte für WTD pro DSCP oder CoS definiert.

In der nächsten Ausgabe werden nur zwei Schwellenwerte definiert.

Wenden Sie zwei Warteschlangenlimits auf eine Richtlinienzuordnung an:

```
<#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
```

Beobachten Sie die Ergebnisse der Pufferzuweisung:

```
<#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  Glb1SMin  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>
```

Im nächsten Beispiel wird eine dritte Warteschlangenlimit-Konfiguration zur Klassenstandardklasse hinzugefügt.

Beobachten Sie die Ergebnisse der Pufferzuweisung:

```
<#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>
```

Wenn eine dritte Warteschlangenlimit-Konfiguration zu einer Warteschlange hinzugefügt wird, wird der 400%ige Multiplikator für den ausgeblendeten weichen Puffer des Systems für diese Warteschlange deaktiviert. Diese Warteschlange akzeptiert jedoch immer noch einen benutzerdefinierten QoS-Warteschlangen-Softmax-Multiplikator <Prozent>.

## Zugehörige Informationen

- [Cisco Catalyst 9000 Switching-Plattformen: Whitepaper zu QoS und Warteschlangen](#)
- [Technischer Support und Downloads von Cisco](#)

## Informationen zu dieser Übersetzung

Cisco hat dieses Dokument maschinell übersetzen und von einem menschlichen Übersetzer editieren und korrigieren lassen, um unseren Benutzern auf der ganzen Welt Support-Inhalte in ihrer eigenen Sprache zu bieten. Bitte beachten Sie, dass selbst die beste maschinelle Übersetzung nicht so genau ist wie eine von einem professionellen Übersetzer angefertigte. Cisco Systems, Inc. übernimmt keine Haftung für die Richtigkeit dieser Übersetzungen und empfiehlt, immer das englische Originaldokument (siehe bereitgestellter Link) heranzuziehen.