

QoS-Planung und -Warteschlangenverwaltung auf den Catalyst Switches der Serie 3550

Inhalt

[Einführung](#)

[Voraussetzungen](#)

[Anforderungen](#)

[Verwendete Komponenten](#)

[Konventionen](#)

[Ausgangswarteschlangen-Funktion für Ports auf Catalyst Switches der Serie 3550](#)

[Funktionen, die sowohl Gigabit- als auch Nicht-Gigabit-Ports unterstützen](#)

[Funktionen, die nur Gigabit-Ports unterstützen](#)

[Funktionen, die nur Nicht-Gigabit-Ports unterstützen](#)

[CoS-to-Queue-Zuordnung](#)

[Warteschlange mit strikter Priorität](#)

[Weighted Round Robin auf Catalyst 3550](#)

[WRED auf Catalyst Switches der Serie 3550](#)

[Einzelfall bei Catalyst Switches der Serie 3550](#)

[Konfiguration der Warteschlangengröße auf Gigabit-Ports](#)

[Warteschlangenverwaltung und Warteschlangengröße auf Nicht-Gigabit-Ports](#)

[Schlussfolgerung](#)

[Zugehörige Informationen](#)

[Einführung](#)

Die Ausgabeplanung stellt sicher, dass bei starker Überbelegung am Ausgang einer Schnittstelle kein wichtiger Datenverkehr verworfen wird. In diesem Dokument werden alle Techniken und Algorithmen erläutert, die bei der Ausgabeplanung für den Cisco Catalyst 3550-Switch zum Einsatz kommen. In diesem Dokument wird auch erläutert, wie die Ausgabeplanung für die Catalyst 3550-Switches konfiguriert und verifiziert wird.

[Voraussetzungen](#)

[Anforderungen](#)

Für dieses Dokument bestehen keine speziellen Anforderungen.

[Verwendete Komponenten](#)

Die Informationen in diesem Dokument basieren auf dem Catalyst 3550, auf dem Cisco IOS®

Software Release 12.1(12c)EA1 ausgeführt wird.

Die Informationen in diesem Dokument wurden von den Geräten in einer bestimmten Laborumgebung erstellt. Alle in diesem Dokument verwendeten Geräte haben mit einer leeren (Standard-)Konfiguration begonnen. Wenn Ihr Netzwerk in Betrieb ist, stellen Sie sicher, dass Sie die potenziellen Auswirkungen eines Befehls verstehen.

Konventionen

Weitere Informationen zu Dokumentkonventionen finden Sie unter [Cisco Technical Tips Conventions](#) (Technische Tipps zu Konventionen von Cisco).

Ausgangswarteschlangen-Funktion für Ports auf Catalyst Switches der Serie 3550

Es gibt zwei Port-Typen für 3550-Switches:

- Gigabit-Ports
- Nicht-Gigabit-Ports (10/100-Mbit/s-Port)

Diese beiden Ports verfügen über unterschiedliche Funktionen. Im verbleibenden Teil dieses Abschnitts werden diese Funktionen zusammengefasst. In den anderen Abschnitten dieses Dokuments werden die Funktionen genauer erläutert.

Funktionen, die sowohl Gigabit- als auch Nicht-Gigabit-Ports unterstützen

Jeder Port des 3550 verfügt über vier verschiedene Ausgabewarteschlangen. Sie können eine dieser Warteschlangen als reine Prioritätswarteschlange konfigurieren. Alle verbleibenden Warteschlangen werden als Warteschlangen mit nicht strikter Priorität konfiguriert und mithilfe von Weighted Round Robin (WRR) verwaltet. Auf allen Ports wird das Paket einer der vier möglichen Warteschlangen auf Basis der Class of Service (CoS) zugewiesen.

Funktionen, die nur Gigabit-Ports unterstützen

Gigabit-Ports unterstützen auch einen Warteschlangenmanagementmechanismus in jeder Warteschlange. Sie können jede Warteschlange so konfigurieren, dass sie entweder WRED (Weighted Random Early Detection) oder Tail Drop mit zwei Schwellenwerten verwendet. Außerdem können Sie die Größe der einzelnen Warteschlangen anpassen (der Puffer, der den einzelnen Warteschlangen zugewiesen ist).

Funktionen, die nur Nicht-Gigabit-Ports unterstützen

Nicht-Gigabit-Ports verfügen über keine Warteschlangenmechanismen wie WRED oder Tail Drop mit zwei Schwellenwerten. Es wird nur FIFO-Warteschlangenverwaltung auf einem 10/100-Mbit/s-Port unterstützt. Sie können die Größe der vier Warteschlangen an diesen Ports nicht ändern. Sie können jedoch pro Warteschlange eine Mindestreservegröße (min.) zuweisen.

CoS-to-Queue-Zuordnung

In diesem Abschnitt wird erläutert, wie der 3550 jedes Paket in eine Warteschlange stellt. Das Paket wird auf Basis der CoS in die Warteschlange gestellt. Jeder der acht möglichen CoS-Werte wird mithilfe des Schnittstellenbefehls `CoS-to-queue` einer der vier möglichen Warteschlangen zugeordnet. Dieses Beispiel zeigt Folgendes:

```
(config-if)#wrr-queue cos-map queue-id cos1... cos8
```

Hier ein Beispiel:

```
3550(config-if)#wrr-queue cos-map 1 0 1
3550(config-if)#wrr-queue cos-map 2 2 3
3550(config-if)#wrr-queue cos-map 3 4 5
3550(config-if)#wrr-queue cos-map 4 6 7
```

In diesem Beispiel werden folgende Punkte eingefügt:

- CoS 0 und 1 in Warteschlange 1 (Q1)
- CoS 2 und 3 in Q2
- CoS 4 und 5 in Q3
- CoS 6 und 7 in Q4

Mit diesem Befehl können Sie die CoS-to-queue-Zuordnung eines Ports überprüfen:

```
cat3550#show mls qos interface gigabitethernet0/1 queueing
GigabitEthernet0/1
...Cos-queue map:
cos-qid
0 - 1
1 - 1
2 - 2
3 - 2
4 - 3
5 - 3
6 - 4
7 - 4...
```

[Warteschlange mit strikter Priorität](#)

Eine Warteschlange mit strikter Priorität wird immer zuerst geleert. Sobald also ein Paket in der Warteschlange mit strikter Priorität vorhanden ist, wird das Paket weitergeleitet. Nachdem jedes Paket von einer der WRR-Warteschlangen weitergeleitet wurde, wird die Warteschlange mit der strikten Priorität überprüft und bei Bedarf geleert.

Eine Warteschlange mit strikter Priorität wurde speziell für verzögerungs-/jitterempfindlichen Datenverkehr wie Sprache entwickelt. Eine strikte Prioritätswarteschlange kann schließlich zum Verhungern der anderen Warteschlangen führen. Pakete, die in den drei anderen WRR-Warteschlangen eingefügt werden, werden nie weitergeleitet, wenn ein Paket in der Warteschlange mit strikter Priorität wartet.

[Tipps](#)

Um zu verhindern, dass die anderen Warteschlangen verhungern, sollten Sie besonders darauf

achten, welcher Datenverkehr in die Prioritätswarteschlange gestellt wird. Diese Warteschlange wird in der Regel für Sprachdatenverkehr verwendet, dessen Volumen in der Regel nicht sehr hoch ist. Wenn jedoch jemand in der Lage ist, großen Datenverkehr mit CoS-Priorität an die Warteschlange mit strikter Priorität (z. B. große Dateiübertragungen oder Backups) zu senden, kann der Ausfall des anderen Datenverkehrs zu Ausfällen führen. Um dieses Problem zu vermeiden, muss besonderer Datenverkehr in die Klassifizierung/Zulassung und Kennzeichnung des Datenverkehrs im Netzwerk aufgenommen werden. Sie können z. B. folgende Vorsichtsmaßnahmen treffen:

- Verwenden Sie den QoS-Status nicht vertrauenswürdiger Ports für alle nicht vertrauenswürdigen Quell-Ports.
- Verwenden Sie die Funktion zur vertrauenswürdigen Begrenzung des Cisco IP-Telefon-Ports, um sicherzustellen, dass dieser nicht im Vertrauenszustand verwendet wird, der für ein IP-Telefon für eine andere Anwendung konfiguriert ist.
- Überwachen Sie den Datenverkehr, der an die Warteschlange mit höchster Priorität weitergeleitet wird. Legen Sie ein Limit für den Richtlinienverkehr mit einem CoS von 5 (Differentiated Services Code Point [DSCP] 46) auf 100 MB an einem Gigabit-Port fest.

Weitere Informationen zu diesen Themen finden Sie in den folgenden Dokumenten:

- [Verständnis von QoS-Richtlinienvergabe und -Kennzeichnung auf dem Catalyst 3550](#)
- [Konfigurieren einer vertrauenswürdigen Grenze zum Sicherstellen der Port-Sicherheit im Abschnitt QoS-Konfiguration](#) (Catalyst 3500)

Auf dem 3550 können Sie eine Warteschlange als Prioritätswarteschlange konfigurieren (immer Q4). Verwenden Sie diesen Befehl im Schnittstellenmodus:

```
3550(config-if)#priority-queue out
```

Wenn die Prioritätswarteschlange nicht in einer Schnittstelle konfiguriert ist, gilt Q4 als Standard-WRR-Warteschlange. Im Abschnitt [Weighted Round Robin on Catalyst 3550](#) dieses Dokuments finden Sie weitere Einzelheiten. Sie können überprüfen, ob die Warteschlange mit der strikten Priorität auf einer Schnittstelle konfiguriert ist, wenn Sie denselben Cisco IOS-Befehl ausführen:

```
NifNif#show mls qos interface gigabitethernet0/1 queueing
GigabitEthernet0/1
Egress expedite queue: ena
```

[Weighted Round Robin auf Catalyst 3550](#)

WRR ist ein Mechanismus, der bei der Ausgabeplanung für den 3550 verwendet wird. WRR arbeitet zwischen drei oder vier Warteschlangen (wenn keine Warteschlange mit strikter Priorität besteht). Die Warteschlangen, die im WRR verwendet werden, werden rundherum geleert, und Sie können das Gewicht für jede Warteschlange konfigurieren.

Sie können beispielsweise Gewichtungen so konfigurieren, dass die Warteschlangen anders verarbeitet werden, wie in der folgenden Liste dargestellt:

- Server-WRR Q1: 10 % der Zeit
- Server-WRR Q2: 20 % der Zeit

- Server WRR Q3: 60 % der Zeit
- Server-WRR Q4: 10 % der Zeit

Für jede Warteschlange können Sie diese Befehle im Schnittstellenmodus ausführen, um die vier Gewichtungen (mit einer für jede Warteschlange) zu konfigurieren:

```
(config-f)#wrr-queue bandwidth weight1 weight2 weight3 weight4
```

Hier ein Beispiel:

```
3550(config)#interface gigabitethernet 0/1
3550(config-if)#wrr-queue bandwidth 1 2 3 4
```

Hinweis: Die Gewichtungen sind relativ. Diese Werte werden verwendet:

- $Q1 = \text{Gewicht 1} / (\text{Gewicht1} + \text{Gewicht2} + \text{Gewicht3} + \text{Gewicht4}) = 1 / (1+2+3+4) = 1/10$
- Frage 2 = 2/10
- Frage 3 = 3/10
- Frage 4 = 4/10

WRR kann auf zwei Arten implementiert werden:

- **WRR pro Bandbreite:** Jedes Gewicht stellt eine bestimmte Bandbreite dar, die gesendet werden darf. Das Gewicht Q1 darf ungefähr 10 Prozent der Bandbreite ausmachen, das zweite Quartal erhält 20 Prozent der Bandbreite usw. Dieses Schema ist derzeit nur in der Catalyst 6500/6000-Serie implementiert.
- **WRR pro Paket:** Dies ist der Algorithmus, der im 3550-Switch implementiert ist. Jedes Gewicht stellt eine bestimmte Anzahl von Paketen dar, die unabhängig von ihrer Größe gesendet werden.

Da der 3550 WRR pro Paket implementiert, gilt dieses Verhalten für die Konfiguration in diesem Abschnitt:

- Q1 überträgt 1 Paket von 10
- Q2 überträgt 2 von 10 Paketen
- Q3 überträgt 3 von 10 Paketen
- 4. Quartal überträgt 4 von 10 Paketen

Alle zu übertragenden Pakete können die gleiche Größe haben. Sie erreichen immer noch eine erwartete Bandbreitenverteilung zwischen den vier Warteschlangen. Unterscheidet sich die durchschnittliche Paketgröße zwischen den Warteschlangen, hat dies große Auswirkungen auf das, was bei einer Überlastung übertragen und verworfen wird.

Nehmen Sie beispielsweise an, dass nur zwei Flüsse im Switch vorhanden sind. Hypothetisch geht man auch davon aus, dass diese Bedingungen gegeben sind:

- Im zweiten Quartal wird ein Gbit/s kleiner interaktiver Anwendungsdatenverkehr (80-Byte-Frames [B]) mit einer CoS von 3 platziert.
- Im ersten Quartal wird ein Gbit/s an Datenverkehr mit großer Dateiübertragung (1518-B-Frames) und einer CoS von 0 platziert.

Zwei Warteschlangen im Switch werden mit Daten von 1 Gbit/s gesendet.

Beide Datenströme müssen denselben Gigabit-Ausgangsport gemeinsam nutzen. Es wird angenommen, dass zwischen Q1 und Q2 das gleiche Gewicht konfiguriert wird. WRR wird pro Paket angewendet, und die Menge der Daten, die von jeder Warteschlange übertragen werden, variiert zwischen den beiden Warteschlangen. Die gleiche Anzahl von Paketen wird aus jeder Warteschlange weitergeleitet, aber der Switch sendet tatsächlich diese Datenmenge:

- 77700 Pakete pro Sekunde (pps) von Q2 = (7700 x 8 x 64) Bit pro Sekunde (bps) (ca. 52 Mbit/s)
- 7700 pps von Q1 = (7700 x 8 x 1500) bps (ca. 948 Mbit/s)

Tipps

- Wenn Sie einen fairen Zugriff für jede Warteschlange auf das Netzwerk zulassen möchten, berücksichtigen Sie die durchschnittliche Größe jedes Pakets. Es wird erwartet, dass jedes Paket in eine Warteschlange gestellt und das Gewicht entsprechend geändert wird. Wenn Sie z. B. gleichen Zugriff auf jede der vier Warteschlangen gewähren möchten, sodass jede Warteschlange 1/4 der Bandbreite erhält, wird der Datenverkehr wie folgt berechnet: In Q1: Bestmöglicher Internetdatenverkehr Angenommen, der Datenverkehr hat eine durchschnittliche Paketgröße von 256 B. Im 2. Quartal: Die Sicherung besteht aus Dateiübertragung, mit einem Paket hauptsächlich 1500 B. Im 3. Quartal: Video-Streams, die auf Paketen mit 192 B erfolgen. Im 4. Quartal: Interaktive Anwendung, die hauptsächlich aus einem Paket von 64 B besteht. Dadurch entstehen folgende Bedingungen: Q1 verbraucht das Vierfache der Bandbreite im 4. Quartal. Q2 verbraucht das 24-fache der Bandbreite im 4. Quartal. Q3 verbraucht dreimal so viel Bandbreite wie Q4.
- Um gleichen Bandbreitenzugriff auf das Netzwerk zu erhalten, konfigurieren Sie Folgendes: Q1 mit einem Gewicht von 6 Q2 mit einem Gewicht von 1 Q3 mit einem Gewicht von 8 Q4 mit einem Gewicht von 24
- Wenn Sie diese Gewichte zuweisen, erreichen Sie eine gleichmäßige Bandbreitenverteilung zwischen den vier Warteschlangen im Fall einer Überlastung.
- Wenn die Warteschlange mit strikter Priorität aktiviert ist, werden die WRR-Gewichtungen auf die drei verbleibenden Warteschlangen umverteilt. Wenn die Warteschlange mit strikter Priorität aktiviert ist und Q4 nicht konfiguriert ist, ist das erste Beispiel mit den Gewichtungen 1, 2, 3 und 4: Q1 = 1 / (1+2+3) = 1 Paket von 6 Q2 = 2 von 6 Paketen Q3 = 3 von 6 Paketen Sie können den Befehl **show** der Cisco IOS-Software ausführen, um das Gewicht der Warteschlange zu überprüfen:

```
NifNif#show mls qos interface gigabitethernet0/1 queueing
GigabitEthernet0/1
QoS is disabled. Only one queue is used
When QoS is enabled, following settings will be applied
Egress expedite queue: dis
wrr bandwidth weights:
qid-weights
 1 - 25
 2 - 25
 3 - 25
 4 - 25
```

Wenn die Prioritätswarteschlange "Expresste Priority" (Beschleunigte Prioritätswarteschlange) aktiviert ist, wird die Gewichtung im 4. Quartal nur verwendet, wenn die Beschleunigungswarteschlange deaktiviert wird. Hier ein Beispiel:

```
NifNif#show mls qos interface gigabitethernet0/1 queueing
```

```
GigabitEthernet0/1
Egress expedite queue: ena
wrr bandwidth weights:
qid-weights
 1 - 25
 2 - 25
 3 - 25
 4 - 25
!--- The expedite queue is disabled.
```

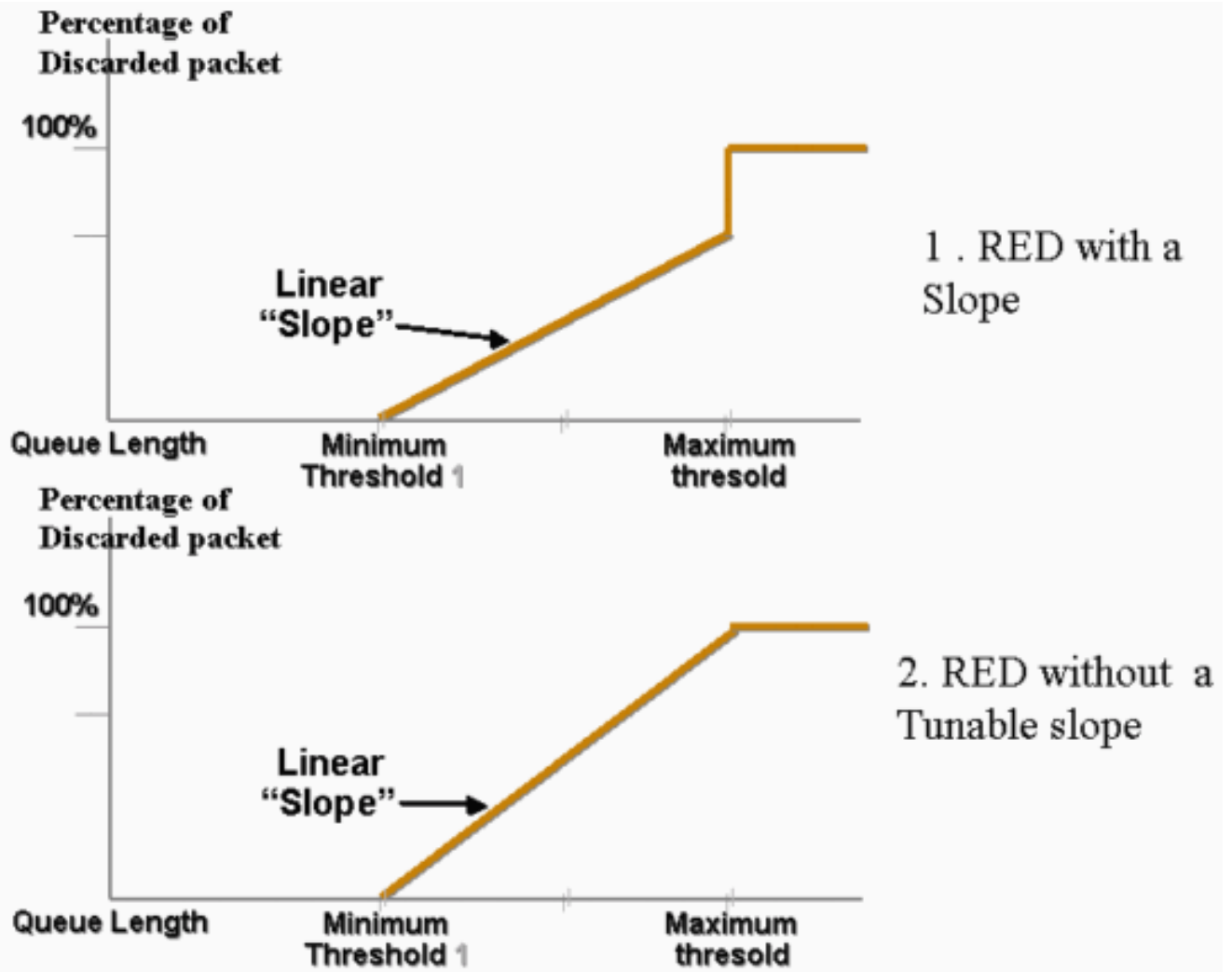
WRED auf Catalyst Switches der Serie 3550

WRED ist nur für Gigabit-Ports auf den Switches der Serie 3550 verfügbar. WRED ist eine Modifikation von Random Early Detection (RED), die zur Überlastungsvermeidung verwendet wird. Für ROT sind folgende Parameter definiert:

- **Mindestschwelle:** Stellt einen Schwellenwert in einer Warteschlange dar. Unter diesem Schwellenwert werden keine Pakete verworfen.
- **Maximaler (max.) Grenzwert:** Stellt einen anderen Schwellenwert in einer Warteschlange dar. Alle Pakete werden oberhalb des maximalen Schwellenwerts verworfen.
- **Neigung:** Möglichkeit zum Verwerfen des Pakets zwischen Min. und Max. Die Fallwahrscheinlichkeit steigt linear (mit einer bestimmten Neigung) mit der Warteschlangengröße.

Dieses Diagramm zeigt die Drop-Wahrscheinlichkeit eines Pakets in der RED-Warteschlange:

Hinweis: Alle Catalyst Switches, die ROT implementieren, können die Neigung anpassen.

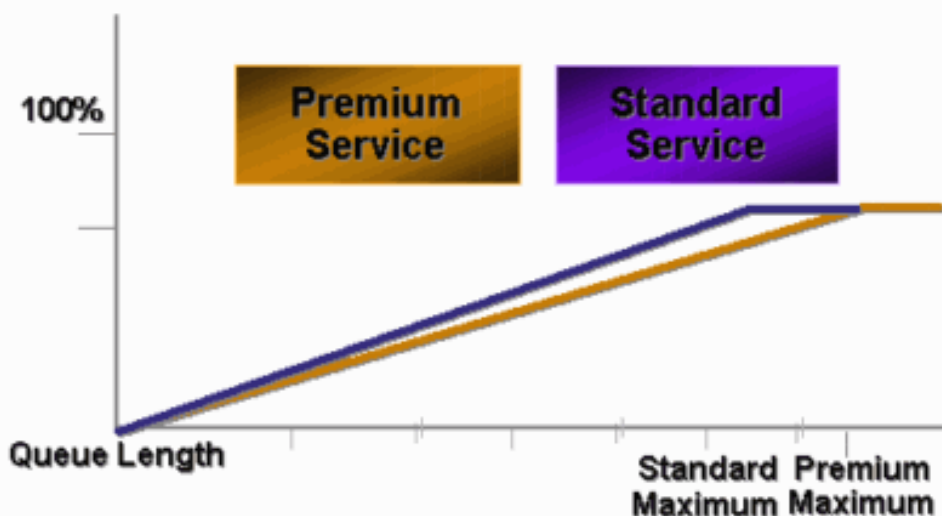


In WRED werden verschiedene Services gewichtet. Sie können einen Standarddienst und einen Premium-Service definieren. Jeder Service erhält einen anderen Satz von Schwellenwerten. Nur Pakete, die dem Standarddienst zugewiesen sind, werden verworfen, wenn der Mindestwert 1 erreicht ist. Nur Pakete von Premium-Services werden verworfen, wenn der Mindestwert 2 erreicht ist. Wenn der Min.-Grenzwert 2 über dem Min.-Schwellenwert 1 liegt, werden mehr Pakete vom Standarddienst verworfen als Pakete von den Premium-Services. Dieses Diagramm zeigt ein Beispiel für die Drop-Wahrscheinlichkeit für jeden Service mit WRED:

Hinweis: Der Switch 3550 ermöglicht es Ihnen nicht, den Mindestschwellenwert, sondern nur den maximalen Schwellenwert einzustellen. Der Mindestwert ist immer fest auf 0 festgelegt. Dadurch ergibt sich eine Fallwahrscheinlichkeit, die das darstellt, was derzeit im 3550 implementiert wird.



3. WRED with 2 Set of min-max threshold (2 services)



4. WRED with 2 Set of service but Min-threshold = 0

Jede Warteschlange, die für WRED auf dem 3550 aktiviert ist, hat immer eine Nicht-Nullpunkt-Wahrscheinlichkeit und verwirft immer Pakete. Dies ist der Fall, da der Mindestwert immer 0 ist. Wenn Sie Paketverluste bei max. vermeiden möchten, verwenden Sie gewichtete Tail Drop, wie im Abschnitt [Tail Drop für Catalyst 3550-Switches](#) beschrieben wird.

Tipp: Die Cisco Bug-ID [CSCdz73556](#) (nur [registrierte](#) Kunden) dokumentiert eine Verbesserungsanfrage für die Konfiguration eines Mindestwerts.

Weitere Informationen zu ROT und WRED finden Sie unter [Übersicht über die Überlastungsvermeidung](#).

Auf dem 3550 können Sie WRED mit zwei verschiedenen maximalen Schwellenwerten konfigurieren, um zwei verschiedene Services bereitzustellen. Je nach Schwellenwert werden unterschiedliche Datenverkehrsarten zugewiesen, die nur von den internen DSCPs abhängen. Dies unterscheidet sich von der Warteschlangenzuweisung, die nur von der CoS des Pakets abhängt. Eine DSCP-zu-Schwellenwert-Tabellenzuordnung bestimmt, bis zu welchem Grenzwert die 64 DSCPs festgelegt werden. Sie können diesen Befehl ausgeben, um diese Tabelle anzuzeigen und zu ändern:

```
(config-if)#wrr-queue dscp-map threshold_number DSCP_1 DSCP_2 DSCP_8
```

Mit diesem Befehl wird beispielsweise DSCP 26 dem Schwellenwert 2 zugewiesen:

```
NifNif(config-if)#wrr-queue dscp-map 2 26
NifNif#show mls qos interface gigabitethernet0/1 queueing
GigabitEthernet0/1
Dscp-threshold map:
  dl :  d2 0  1  2  3  4  5  6  7  8  9
-----
  0 :    01 01 01 01 01 01 01 01 01 01 01
  1 :    01 01 01 01 01 01 01 02 01 01 01
  2 :    01 01 01 01 02 01 02 01 01 01 01
  3 :    01 01 01 01 01 01 01 01 01 01 01
  4 :    02 01 01 01 01 01 02 01 01 01 01
  5 :    01 01 01 01 01 01 01 01 01 01 01
  6 :    01 01 01 01
```

Nach der Definition der DSCP-zu-Schwellenwert-Zuordnung ist WRED in der Warteschlange Ihrer Wahl aktiviert. Geben Sie den folgenden Befehl ein:

```
(config-if)#wrr-queue random-detect max-threshold queue_id threshold_1 threshold_2
```

In diesem Beispiel wird Folgendes konfiguriert:

- Q1 mit Schwellenwert 1 = 50 % und Schwellenwert 2 = 100 %
- Q2 mit Schwellenwert 1 = 70 % und Schwellenwert 2 = 100 %

```
3550(config)#interface gigabitethernet 0/1
3550(config-if)#wrr-queue random-detect max-threshold 1 50 100
3550(config-if)#wrr-queue random-detect max-threshold 2 70 100
3550(config-if)#wrr-queue random-detect max-threshold 3 50 100
3550(config-if)#wrr-queue random-detect max-threshold 4 70 100
```

Sie können diesen Befehl ausführen, um den Warteschlangentyp (WRED oder nicht) für jede Warteschlange zu überprüfen:

```
nifnif#show mls qos interface gigabitethernet0/1 buffers
GigabitEthernet0/1
..
qid WRED thresh1 thresh2
1 dis 10 100
2 dis 10 100
3 ena 10 100
4 dis 100 100
```

Die `ena` bedeutet enable, und die Warteschlange verwendet WRED. Das `dis` bedeutet "disable", und die Warteschlange verwendet Tail Drop.

Sie können auch die Anzahl der Pakete überwachen, die für jeden Grenzwert verworfen werden. Geben Sie den folgenden Befehl ein:

```
show mls qos interface gigabitethernetx/x statistics
WRED drop counts:
qid thresh1 thresh2 FreeQ
1 : 327186552 8 1024
2 : 0 0 1024
3 : 37896030 0 1024
```

[Einzelfall bei Catalyst Switches der Serie 3550](#)

Tail Drop ist der Standardmechanismus auf den 3550-Ports auf Gigabit-Ports. Jeder Gigabit-Port kann zwei Tail Drop-Schwellenwerte haben. Jedem Tail-Drop-Grenzwert werden eine Reihe von DSCPs zugewiesen, wobei dieselbe DSCP-Schwellenwertzuordnung verwendet wird, die auch der Abschnitt [WRED auf Catalyst 3550-Switches](#) in diesem Dokument definiert. Wenn ein Schwellenwert erreicht ist, werden alle Pakete mit einem DSCP, dem dieser Grenzwert zugewiesen ist, verworfen. Sie können diesen Befehl ausführen, um SchwanzDrop-Schwellenwerte zu konfigurieren:

```
(config-if)#wrr-queue threshold queue-id threshold-percentage1 threshold-percentage2
```

In diesem Beispiel wird Folgendes konfiguriert:

- Q1 mit Schwanzschwellenwert 1 = 50 % und Schwellenwert 2 = 100 %
- Q2 mit Schwellenwert 1 = 70 % und Schwellenwert 2 = 100 %

```
Switch(config-if)#wrr-queue threshold 1 50 100  
Switch(config-if)#wrr-queue threshold 2 70 100  
Switch(config-if)#wrr-queue threshold 3 60 100  
Switch(config-if)#wrr-queue threshold 4 80 100
```

[Konfiguration der Warteschlangengröße auf Gigabit-Ports](#)

Der Switch 3550 verwendet eine zentrale Pufferung. Das bedeutet, dass es keine festen Puffergrößen pro Port gibt. Es gibt jedoch eine feste Anzahl von Paketen auf einem Gigabit-Port, die in die Warteschlange gestellt werden können. Diese feste Nummer ist 4096. Standardmäßig kann jede Warteschlange in einem Gigabit-Port unabhängig von der Paketgröße bis zu 1024 Pakete enthalten. Sie können jedoch die Aufteilung dieser 4096-Pakete auf die vier Warteschlangen ändern. Geben Sie den folgenden Befehl ein:

```
wrr-queue queue-limit Q_size1 Q_size2 Q_size3 Q_size4
```

Hier ein Beispiel:

```
3550(config)#interface gigabitethernet 0/1  
3550(config-if)#wrr-queue queue-limit 4 3 2 1
```

Diese Parameter für die Warteschlangengröße sind relativ. Dieses Beispiel zeigt Folgendes:

- Q1 ist viermal größer als Q4.
- Q2 ist dreimal größer als Q4.
- Q3 ist doppelt so groß wie Q4.

Die 4096-Pakete werden folgendermaßen umverteilt:

- $Q1 = [4 / (1+2+3+4)] * 4096 = 1639$ Pakete
- $Q2 = 0,3 * 4096 = 1229$ Pakete
- $Q3 = 0,2 * 4096 = 819$ Pakete
- $Q4 = 0,1 * 4096 = 409$ Pakete

Mit diesem Befehl können Sie die relativen Gewichte von geteilten Puffern in den vier Warteschlangen anzeigen:

```
cat3550#show mls qos interface buffers
GigabitEthernet0/1
Notify Q depth:
qid-size
1 - 4
2 - 3
3 - 2
4 - 1
...
```

Sie können diesen Befehl auch ausführen, um zu sehen, wie viele freie Pakete pro Warteschlange noch enthalten sein können:

```
(config-if)#show mls qos interface gigabitethernetx/x statistics
WRED drop counts:
qid thresh1 thresh2 FreeQ
1 : 0      0      1639
2 : 0      0      1229
3 : 0      0      819
4 : 0      0      409
```

Der `FreeQ` count Parameter ist dynamisch. Der `FreeQ`-Zähler gibt die maximale Warteschlangengröße abzüglich der Anzahl der Pakete an, die sich derzeit in der Warteschlange befinden. Wenn es im 1. Quartal derzeit 39 Pakete gibt, sind in der `FreeQ`-Anzahl 1600 Pakete frei. Hier ein Beispiel:

```
(config-if)#show mls qos interface gigabitethernetx/x statistics
WRED drop counts:
qid thresh1 thresh2 FreeQ
1 : 0      0      1600
2 : 0      0      1229
3 : 0      0      819
4 : 0      0      409
```

Warteschlangenverwaltung und Warteschlangengröße auf Nicht-Gigabit-Ports

Für 10/100-Mbit/s-Ports ist kein Warteschlangenmanagementschema verfügbar (kein WRED oder Tail Drop mit zwei Schwellenwerten). Alle vier Warteschlangen sind FIFO-Warteschlangen. Außerdem gibt es keine maximale Warteschlangengröße, die 4096 Pakete für jeden Gigabit-Port reserviert. 10/100-Mbit/s-Ports speichern Pakete in jeder Warteschlange, bis sie aufgrund fehlender Ressourcen voll sind. Sie können eine Mindestanzahl von Paketen pro Warteschlange reservieren. Dieser Mindestwert ist standardmäßig auf 100 Pakete pro Warteschlange festgelegt. Sie können diesen Mindestreservewert für jede Warteschlange ändern, wenn Sie unterschiedliche Mindestreservewerte definieren und jeder Warteschlange einen Wert zuweisen.

Gehen Sie wie folgt vor, um diese Änderung vorzunehmen:

1. Weisen Sie für jeden globalen Mindestreservewert eine Puffergröße zu. Sie können maximal acht verschiedene Mindestreservewerte konfigurieren. Geben Sie den folgenden Befehl ein:

```
(Config)# mls qos min-reserve min-reserve-level min-reserve-buffersize
```

Diese Mindestreservewerte sind global für den Switch. Standardmäßig sind alle Mindestreservewerte auf 100 Pakete festgelegt. Führen Sie zum Beispiel folgende Befehle aus, um eine Mindestreserve-Ebene 1 von 150 Paketen und eine Mindestreserve-Ebene 2 von 50 Paketen zu konfigurieren:

```
nifnif(config)#mls qos min-reserve ?
<1-8> Configure min-reserve level
nifnif(config)#mls qos min-reserve 1 ?
<10-170> Configure min-reserve buffers
nifnif(config)#mls qos min-reserve 1 150
nifnif(config)#mls qos min-reserve 2 50
```

2. Weisen Sie jeder Warteschlange einen der Mindestreservewerte zu. Sie müssen jede Warteschlange einem der Mindestreservewerte zuweisen, um zu wissen, wie viele Puffer für diese Warteschlange garantiert werden. Diese Bedingungen gelten standardmäßig: Q1 wird der Mindestreservepflicht 1 zugewiesen. Q2 wird der Mindestreservepflicht 2 zugewiesen. Q3 wird der Mindestreservestufe 3 zugewiesen. Das 4. Quartal ist der Mindestreservepflicht 4 zugewiesen. Standardmäßig sind alle Mindestreservewerte 100. Sie können diesen Schnittstellenbefehl ausstellen, um pro Warteschlange einen anderen Mindestreservewert zuzuweisen:

```
(config-if)#wrr-queue min-reserve queue-id min-reserve-level
```

Um beispielsweise Q1 eine Mindestreserve von 2 und Q2 eine Mindestreserve von 1 zuzuweisen, führen Sie den folgenden Befehl aus:

```
nifnif(config)#interface fastethernet 0/1
nifnif(config-if)#wrr-queue min-reserve ?
<1-4> queue id
nifnif(config-if)#wrr-queue min-reserve 1 ?
<1-8> min-reserve level
nifnif(config-if)#wrr-queue min-reserve 1 2
nifnif(config-if)#wrr-queue min-reserve 2 1
```

Sie können diesen Befehl ausgeben, um die Mindestreservezuweisung zu überprüfen, die folgende Ergebnisse liefert:

```
nifnif#show mls qos interface fastethernet0/1 buffers
FastEthernet0/1
Minimum reserve buffer size:
150 50 100 100 100 100 100 100
!--- This shows the value of all eight min reserve levels. Minimum reserve buffer level
select: 2 1 3 4 !--- This shows the min reserve level that is assigned to !--- each queue
(from Q1 to Q4).
```

Schlussfolgerung

Die Warteschlangenverwaltung und Planung für einen Port des 3550 umfasst folgende Schritte:

1. Weisen Sie jede CoS einer Warteschlange zu.
2. Aktivieren Sie bei Bedarf Warteschlangen mit strikter Priorität.
3. Weisen Sie das WRR-Gewicht zu, und berücksichtigen Sie die erwartete Paketgröße in der Warteschlange.

4. Ändern Sie die Warteschlangengröße (nur Gigabit-Ports).
5. Aktivieren Sie einen Warteschlangenmanagementmechanismus (Tail Drop oder WRED, nur für Gigabit-Ports).

Durch eine ordnungsgemäße Warteschlangenverwaltung und Planung können Verzögerungen/Jitter für Sprach-/Videodatenverkehr reduziert und Verluste für geschäftskritischen Datenverkehr vermieden werden. Beachten Sie die folgenden Richtlinien für eine maximale Scheduling-Leistung:

- Klassifizieren Sie den im Netzwerk vorhandenen Datenverkehr in verschiedenen Klassen, entweder mit vertrauenswürdiger oder spezifischer Markierung.
- Übermäßiger Polizeiverkehr.

Zugehörige Informationen

- [Verständnis von QoS-Richtlinienvergabe und -Kennzeichnung auf dem Catalyst 3550](#)
- [Konfigurieren von QoS - Produktdokumentation](#)
- [Support-Seiten für LAN-Produkte](#)
- [Support-Seite für LAN-Switching](#)
- [Technischer Support und Dokumentation - Cisco Systems](#)