Domande frequenti sulla funzionalità QoS

Sommario

Introduzione

In questo documento vengono riportate le risposte alle domande frequenti sulla funzionalità QoS (Quality of Service).

Informazioni generali

D. Che cos'è QoS (Quality of Service)?

R. QoS indica la capacità di una rete di fornire un servizio migliore a una selezione del traffico di rete su diverse tecnologie sottostanti, tra cui Frame Relay, ATM (Asynchronous Transfer Mode), reti Ethernet e 802.1, SONET e reti con routing IP.

QoS è una raccolta di tecnologie che consente alle applicazioni di richiedere e ricevere livelli di servizio prevedibili in termini di capacità di throughput dei dati (larghezza di banda), variazioni di latenza (jitter) e ritardo. In particolare, le funzionalità QoS forniscono un servizio di rete migliore e più prevedibile mediante i seguenti metodi:

- Supporto della larghezza di banda dedicata.
- Miglioramento delle caratteristiche di perdita.
- Evitare e gestire la congestione della rete.
- Forma del traffico di rete.
- Impostazione delle priorità del traffico in rete.

L'Internet Engineering Task Force (IETF) definisce le due architetture seguenti per QoS:

- Servizi integrati (IntServ)
- Servizi differenziati (DiffServ)

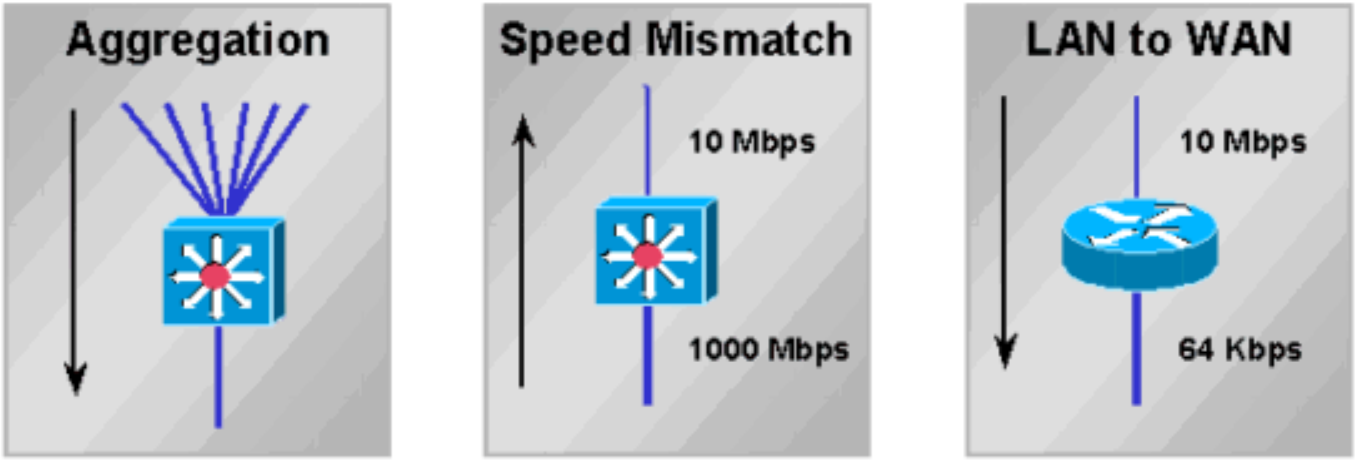
IntServ utilizza il protocollo RSVP (Resource Reservation Protocol) per segnalare esplicitamente le esigenze QoS del traffico di un'applicazione lungo i dispositivi nel percorso end-to-end attraverso la rete. Se ogni dispositivo di rete lungo il percorso è in grado di riservare la larghezza di banda necessaria, l'applicazione di origine può iniziare la trasmissione. Request for Comments (RFC) 2205 definisce RSVP e RFC 1633 definisce IntServ.

DiffServ si concentra sulla QoS aggregata e con provisioning. Anziché segnalare i requisiti QoS di un'applicazione, DiffServ utilizza un DSCP (DiffServ Code Point) nell'intestazione IP per indicare i livelli QoS richiesti. Il software Cisco IOS® versione 12.1(5)T ha introdotto la conformità DiffServ sui router Cisco. Per ulteriori informazioni, consultare i seguenti documenti:

- Servizio integrato in Cisco IOS 12.1
- Implementazione di DiffServ per una qualità del servizio completa
- Implementazione delle policy della qualità del servizio (QoS) con DSCP

D. Cosa sono congestione, ritardo e jitter?

R. Un'interfaccia risulta congestionata quando riceve più traffico di quello che è in grado di gestire. I punti di congestione della rete sono candidati affidabili per i meccanismi QoS (Quality of Service). Di seguito è riportato un esempio di punti di congestione tipici:



La congestione della rete comporta un ritardo. Una rete e i suoi dispositivi introducono diversi tipi di ritardi, come spiegato in Informazioni sul ritardo nelle reti voce dei pacchetti. La variazione di ritardo è nota come jitter, come spiegato in Descrizione della jitter nelle reti voce pacchetti (piattaforme Cisco IOS). Per supportare il traffico interattivo e in tempo reale, è necessario controllare e ridurre al minimo i ritardi e gli jitter.

D. Che cos'è MQC?

A. MQC è l'acronimo di Modular Quality of Service (QoS) Command Line Interface (CLI). È stato progettato per semplificare la configurazione della funzionalità QoS sui router e sugli switch Cisco, definendo una sintassi di comando comune e il risultante set di comportamenti QoS sulle piattaforme. Questo modello sostituisce il modello precedente di definizione di sintassi univoca per ciascuna funzionalità QoS e per ciascuna piattaforma.

Il processo MQC prevede i tre passaggi seguenti:

1. Definire una classe di traffico utilizzando il comando class-map.
2. Creare un criterio del traffico associando la classe del traffico a una o più funzionalità QoS mediante il comando policy-map.
3. Collegare i criteri del traffico all'interfaccia, alla sottointerfaccia o al circuito virtuale (VC) usando il comando service-policy.

Nota: è possibile implementare le funzioni di condizionamento del traffico di DiffServ, ad esempio la marcatura e il shaping, utilizzando la sintassi MQC.

Per ulteriori informazioni, consultare il documento sull'interfaccia della riga di comando Modular Quality of Service.

D. Qual è il significato `dei criteri di servizio supportati solo sulle interfacce VIP con` messaggio `abilitato per` DCEF?

R. Sui Versatile Interface Processor (VIP) di un Cisco serie 7500, le funzionalità QoS (Quality of Service) distribuite sono supportate solo a partire dalla versione Cisco IOS 12.1(5)T, 12.1(5)E e 12.0(14)S. L'abilitazione di Cisco Express Forwarding (dCEF) distribuito abilita automaticamente QoS distribuito.

Le interfacce non VIP, note come IP (legacy Interface Processor), supportano le funzionalità QoS centrali come abilitate sul Route Switch Processor (RSP). Per ulteriori informazioni, consultare i seguenti documenti:

- Distributed Class-Based Weighted Fair Queueing e Distributed Weighted Early Detection
- Accodamento a bassa latenza distribuita
- Distributed Traffic Shaping
- Versatile Interface Processor-Based Distributed FRF.11 e FRF.12 per Cisco IOS release 12.1 T

D. Quante classi sono supportate da una policy QoS (Quality of Service)?

R. Nelle versioni di Cisco IOS precedenti alla 12.2 era possibile definire un massimo di 256 classi e un massimo di 256 classi all'interno di ciascun criterio, se le stesse classi vengono riutilizzate per criteri diversi. Se si dispone di due criteri, il numero totale di classi di entrambi i criteri non deve superare 256. Se un criterio include CBWFQ (Class-Based Weighted Fair Queueing), ovvero

include un'istruzione di larghezza di banda [o priorità] all'interno di una delle classi, il numero totale di classi supportate è 64.

Nelle versioni 12.2(12), 12.2(12)T e 12.2(12)S di Cisco IOS, questa limitazione di 256 mappe di classi globali è stata modificata ed è ora possibile configurare fino a 1024 mappe di classi globali e utilizzare 256 mappe di classi all'interno della stessa mappa di criteri.

D. In che modo vengono elaborati gli aggiornamenti di routing e i pacchetti keepalive PPP (Point-to-Point Protocol) / HDLC (High-Level Data Link Control) quando viene applicata una policy sui servizi?

R. I router Cisco IOS utilizzano i due meccanismi seguenti per assegnare la priorità ai pacchetti di controllo:

- Precedenza IP
- priorità_pak

Entrambi i meccanismi sono progettati per garantire che i pacchetti di controllo della chiave non vengano scartati per ultimi dal router e dal sistema di coda quando un'interfaccia in uscita è congestionata. Per ulteriori informazioni, consultare il documento sulla modalità di accodamento degli aggiornamenti del routing e dei pacchetti di controllo su un'interfaccia con un criterio del servizio QoS.

D. La funzionalità QoS (Quality of Service) è supportata sulle interfacce configurate con IRB (Integrated Routing and Bridging)?

R. No. Non è possibile configurare le funzionalità QoS quando l'interfaccia è configurata per IRB.

Classificazione e contrassegno

D. Che cos'è la preclassificazione QoS (Quality of Service)?

R. La preclassificazione QoS consente di individuare e classificare il contenuto dell'intestazione IP originale dei pacchetti da incapsulare e/o crittografare nel tunnel. Questa funzionalità non descrive il processo di copia del valore originale del byte ToS (Type of service) dall'intestazione del pacchetto originale all'intestazione del tunnel. Per ulteriori informazioni, consultare i seguenti documenti:

- Configurazione di QoS per reti private virtuali
- Quality of Service per reti private virtuali, 12.2(2)T Feature Module

D. Quali campi dell'intestazione del pacchetto possono essere segnalati? Quali valori sono disponibili?

R. La funzione di contrassegno basata su classi consente di impostare o contrassegnare

l'intestazione layer 2, layer 3 o Multiprotocol Label Switching (MPLS) dei pacchetti. Per ulteriori informazioni, consultare i seguenti documenti:

- Configurazione del contrassegno pacchetti basato su classi
- Quando il bit CLP è impostato da un router in una cella ATM?
- Configurazione del contrassegno pacchetti sui PVC Frame Relay

D. È possibile assegnare la priorità al traffico in base all'URL?

R. Sì. NBAR (Network Based Application Recognition) consente di classificare i pacchetti in base ai campi corrispondenti a livello di applicazione. Prima dell'introduzione di NBAR, la classificazione più granulare era quella dei numeri di porta TCP (Transmission Control Protocol) di livello 4 e UDP (User Datagram Protocol). Per ulteriori informazioni, consultare i seguenti documenti:

- Domande e risposte sul riconoscimento delle applicazioni in rete
- Reti di applicazioni NBAR
- Utilizzo di elenchi di riconoscimento delle applicazioni e controllo degli accessi basati sulla rete per il blocco del codice Red Worm
- Come proteggere la rete dal virus Nimda

D. Quali piattaforme e versioni dei software Cisco IOS supportano NBAR (Network Based Application Recognition)?

R. Il supporto per NBAR è stato introdotto nelle seguenti versioni del software Cisco IOS:

Piattaforma	Versione minima del software Cisco IOS
7200	12.1(5)T
7100	12.1(5)T
3660	12.1(5)T
3640	12.1(5)T
3620	12.1(5)T
2600	12.1(5)T
1700	12.2(2)T

Nota: per utilizzare NBAR, è necessario abilitare Cisco Express Forwarding (CEF).

DNBAR (Distributed NBAR) è disponibile sulle seguenti piattaforme:

Piattaforma	Versione minima del software Cisco IOS
7500	12.2(4)T, 12.1(6)E

FlexWAN	12.1(6)E

Nota: NBAR non è supportato sulle interfacce VLAN di Catalyst 6000 Multilayer Switch Feature Card (MSFC), Cisco serie 12000 o il Route Switch Module (RSM) per Catalyst serie 5000. Se non è presente una particolare piattaforma, contattare il rappresentante tecnico Cisco.

Accodamento e gestione della congestione

D. Qual è lo scopo dell'accodamento?

R. L'accodamento è progettato per gestire la congestione temporanea sull'interfaccia di un dispositivo di rete memorizzando i pacchetti in eccesso nei buffer finché non diventa disponibile larghezza di banda. I router Cisco IOS supportano diversi metodi di coda per soddisfare i diversi requisiti di larghezza di banda, jitter e ritardo delle diverse applicazioni.

Il meccanismo predefinito per la maggior parte delle interfacce è il FIFO (First In First Out). Alcuni tipi di traffico hanno requisiti di ritardo/jitter più severi. Pertanto, uno dei seguenti meccanismi di coda alternativi deve essere configurato o è abilitato per impostazione predefinita:

- WFQ (Weighted Fair Queueing)
- CBWFQ (Class-Based Weighted Fair Queueing)
- LLQ (Low-Latency Queueing), ovvero CBWFQ con una Priority Queue (PQ) (nota come PQCBWFQ)
- Priority Queueing (PQ)
- CQ (Custom Queueing)

L'accodamento in genere viene eseguito solo sulle interfacce in uscita. Un router accoda i pacchetti in uscita da un'interfaccia. È possibile controllare il traffico in entrata, ma in genere non è possibile metterlo in coda (un'eccezione è rappresentata dal buffering sul lato ricezione su un router Cisco serie 7500 che utilizza Cisco Express Forwarding (dCEF) distribuito per inoltrare i pacchetti dall'entrata all'interfaccia di uscita; per ulteriori informazioni, fare riferimento a Informazioni sull'esecuzione della CPU VIP al 99% e sul buffering sul lato ricezione. Sulle piattaforme distribuite di fascia alta, come le serie Cisco 7500 e 12000, l'interfaccia in entrata può utilizzare i propri buffer di pacchetto per archiviare il traffico in eccesso commutato su un'interfaccia in uscita congestionata a seguito della decisione di switching dell'interfaccia in entrata. In rare condizioni, in genere quando l'interfaccia in entrata alimenta un'interfaccia in uscita più lenta, l'interfaccia in entrata può sperimentare errori ignorati in aumento quando la memoria del pacchetto è esaurita. Un'eccessiva congestione può causare perdite nella coda di output. Le interruzioni nella coda di input hanno per la maggior parte del tempo una causa principale diversa. Per ulteriori informazioni sulla risoluzione dei problemi relativi alle perdite di dati, consultare il seguente documento:

- Risoluzione dei problemi relativi ai pacchetti eliminati nelle code di input e di output

Per ulteriori informazioni, consultare i seguenti documenti:

- Risoluzione dei problemi "ignorati" su un ATM Port Adapter
- Risoluzione dei problemi relativi agli errori ignorati e all'assenza di perdite di memoria sul Cisco serie 12000 Internet Router

D. Come funzionano WFQ (Weighted Fair Queueing) e CBWFQ (Weighted Fair Queueing) basati su classi?

R. Una coda equa cerca di assegnare una parte equa della larghezza di banda di un'interfaccia tra le conversazioni attive o i flussi IP. Classifica i pacchetti in code secondarie, identificate da un numero di identificazione della conversazione, utilizzando un algoritmo di hashing basato su diversi campi dell'intestazione IP e sulla lunghezza del pacchetto. Di seguito viene illustrato il modo in cui viene calcolato il peso:

- W=K/(precedenza +1)

K= 4096 con Cisco IOS versione 12.0(4)T e precedenti, e 32384 con Cisco IOS versione 12.0(5)T e successive.

Più basso è il peso, maggiore sarà la priorità e la quota della larghezza di banda. Oltre al peso, si tiene conto anche della lunghezza del pacchetto.

CBWFQ consente di definire una classe di traffico e assegnarle una garanzia di larghezza di banda minima. L'algoritmo alla base di questo meccanismo è WFQ, che ne spiega il nome. Per configurare CBWFQ, è necessario definire classi specifiche nelle istruzioni map-class. Assegnare quindi un criterio a ogni classe in una mappa dei criteri. Questa mappa dei criteri verrà quindi collegata in uscita a un'interfaccia. Per ulteriori informazioni, consultare i seguenti documenti:

- Informazioni su Class Based Weighted Fair Queuing su ATM
- Informazioni su Weighted Fair Queuing su ATM

D. Se una classe in CBWFQ (Class Based Weighted Fair Queueing) non utilizza la larghezza di banda, è possibile che altre classi utilizzino tale larghezza di banda?

R. Sì. Anche se le garanzie di larghezza di banda fornite dall'emissione dei comandi bandwidth e priority sono state descritte con parole come "riservate" e "larghezza di banda da lasciare da parte", nessuno dei due comandi implementa una vera prenotazione. Se una classe di traffico non utilizza la larghezza di banda configurata, la larghezza di banda inutilizzata viene condivisa tra le altre classi.

Il sistema di coda impone un'eccezione importante a questa regola con una classe di priorità. Come accennato in precedenza, il carico offerto di una classe di priorità viene misurato da un sorvegliante del traffico. Durante le condizioni di congestione, una classe di priorità non può utilizzare larghezza di banda in eccesso. Per ulteriori informazioni, fare riferimento a Confronto dei comandi di larghezza di banda e priorità di un criterio del servizio QoS.

D. CBWFQ (Class Based Weighted Fair Queueing) è supportato sulle sottointerfacce?

R. Le interfacce logiche Cisco IOS non supportano per loro natura uno stato di congestione e non supportano l'applicazione diretta di un criterio del servizio che applica un metodo di coda. Al contrario, è necessario applicare il shaping alla sottointerfaccia utilizzando il GTS (Generic Traffic Shaping) o il shaping basato su classi. Per ulteriori informazioni, consultare il documento sull'applicazione delle funzionalità QoS alle sottointerfacce Ethernet.

D. Qual è la differenza tra le istruzioni relative alla `priorità` e alla larghezza di `banda` in una mappa delle politiche?

R. I comandi priority e bandwidth differiscono sia per le funzionalità sia per le applicazioni in cui in genere supportano. La tabella seguente riepiloga queste differenze:

Funzione	Comando larghezza di banda	comando priority
Garanzia larghezza di banda minima	Sì	Sì
Massima garanzia di larghezza di banda	No	Sì
Policer integrato	No	Sì
Fornisce bassa latenza	No	Sì

Per ulteriori informazioni, fare riferimento a Confronto tra i comandi di larghezza di banda e priorità di un criterio del servizio QoS.

D. Come viene calcolato il limite di coda sui processori FlexWAN e Versatile Interface (VIP)?

R. Presupponendo che la SRAM sia sufficiente sul VIP o FlexWAN, il limite della coda viene calcolato in base a un ritardo massimo di 500 ms con dimensioni medie del pacchetto di 250 byte. Di seguito è riportato un esempio di classe con una larghezza di banda di 1 Mbps:

Limite coda = 1000000 / (250 x 8 x 2) = 250

Limiti di coda più piccoli vengono assegnati man mano che diminuisce la quantità di memoria del pacchetto disponibile e con un numero maggiore di circuiti virtuali (VCS).

Nell'esempio seguente, un PA-A3 viene installato in una scheda FlexWAN per Cisco serie 7600 e supporta più sottointerfacce con PVC (Permanent Virtual Circuit) da 2 MB. I criteri del servizio vengono applicati a ogni VC.

<#root>

```
class-map match-any XETRA-CLASS
  match access-group 104
class-map match-any SNA-CLASS
  match access-group 101
  match access-group 102
  match access-group 103
policy-map
```

POLICY-2048Kbps

```
  class XETRA-CLASS
    bandwidth 320
  class SNA-CLASS
    bandwidth 512

interface ATM6/0/0
 no ip address
 no atm sonet ilmi-keepalive
 no ATM ilmi-keepalive
!
interface ATM6/0/0.11 point-to-point
 mtu 1578
 bandwidth 2048
 ip address 22.161.104.101 255.255.255.252
 pvc ABCD
  class-vc 2048Kbps-PVC
  service-policy out
```

POLICY-2048Kbps

All'interfaccia ATM (Asynchronous Transfer Mode) viene assegnato un limite di coda per l'intera interfaccia. Il limite è una funzione del totale di buffer disponibili, del numero di interfacce fisiche sulla FlexWAN e del ritardo massimo di accodamento consentito sull'interfaccia. A ogni PVC viene assegnata una parte del limite dell'interfaccia in base all'SCR (Sustained Cell Rate) o all'MCR (Minimum Cell Rate) del PVC e a ogni classe viene assegnata una parte del limite del PVC in base all'allocazione della larghezza di banda.

L'output di esempio seguente del comando show policy-map interface deriva da una FlexWAN con 3687 buffer globali. Utilizzare il comando show buffer per visualizzare questo valore. A ciascun PVC di due Mbps vengono assegnati 50 pacchetti in base alla larghezza di banda del PVC di due Mbps (2047/149760 x 3687 = 50). A ciascuna classe viene quindi allocata una porzione del 50, come illustrato nell'output seguente:

<#root>

```
service-policy output: POLICY-2048Kbps
    class-map: XETRA-CLASS (match-any)
      687569 packets, 835743045 bytes
      5 minute offered rate 48000 bps, drop rate 6000 BPS
      match: access-group 104
        687569 packets, 835743045 bytes
        5 minute rate 48000 BPS
      queue size 0,
```

```
queue limit 7


      packets output 687668, packet drops 22
      tail/random drops 22, no buffer drops 0, other drops 0
      bandwidth: kbps 320, weight 15

    class-map: SNA-CLASS (match-any)
      2719163 packets, 469699994 bytes
      5 minute offered rate 14000 BPS, drop rate 0 BPS
      match: access-group 101
        1572388 packets, 229528571 bytes
        5 minute rate 14000 BPS
      match: access-group 102
        1146056 packets, 239926212 bytes
        5 minute rate 0 BPS
      match: access-group 103
        718 packets, 245211 bytes
        5 minute rate 0 BPS
      queue size 0,

queue limit 12


      packets output 2719227, packet drops 0
      tail/random drops 0, no buffer drops 0, other drops 0
      bandwidth: kbps 512, weight 25
      queue-limit 100

    class-map: class-default (match-any)
      6526152 packets, 1302263701 bytes
      5 minute offered rate 44000 BPS, drop rate 0 BPS
      match: any
        6526152 packets, 1302263701 bytes
        5 minute rate 44000 BPS
      queue size 0,

queue limit 29


      packets output 6526840, packet drops 259
      tail/random drops 259, no buffer drops 0, other drops 0
```

Se i flussi di traffico utilizzano pacchetti di grandi dimensioni, l'output del comando show policy-map interface potrebbe riportare un valore incrementale per il campo `no buffer drops` poiché è possibile che i buffer si esauriscano prima di raggiungere il limite della coda. In questo caso, provare a regolare manualmente il limite della coda nelle classi non prioritarie. Per ulteriori informazioni, consultare il documento sulla definizione del limite della coda di trasmissione con IP al CoS ATM.

D. Come si verifica il valore limite della coda?

R. Nelle piattaforme non distribuite, il limite predefinito è 64 pacchetti. L'output di esempio seguente è stato acquisito su un router Cisco serie 3600:

```
november#
show policy-map interface s0


   Serial0

   Service-policy output: policy1

     Class-map: class1 (match-all)
       0 packets, 0 bytes
       5 minute offered rate 0 BPS, drop rate 0 BPS
       Match: ip precedence 5
       Weighted Fair Queueing
         Output Queue: Conversation 265
         Bandwidth 30 (kbps) Max Threshold 64 (packets)

!--- Max Threshold is the queue-limit.

         (pkts matched/bytes matched) 0/0
         (depth/total drops/no-buffer drops) 0/0/0

      Class-map: class2 (match-all)
        0 packets, 0 bytes
        5 minute offered rate 0 BPS, drop rate 0 BPS
        Match: ip precedence 2
        Match: ip precedence 3
        Weighted Fair Queueing
          Output Queue: Conversation 266
          Bandwidth 24 (kbps) Max Threshold 64 (packets)
          (pkts matched/bytes matched) 0/0
          (depth/total drops/no-buffer drops) 0/0/0

       Class-map: class-default (match-any)
         0 packets, 0 bytes
         5 minute offered rate 0 BPS, drop rate 0 BPS
         Match: any
```

D. È possibile consentire una coda equa all'interno di una classe?

R. Cisco serie 7500 con QoS (Quality of Service) distribuito supporta code eque per classe. Altre piattaforme, tra cui Cisco serie 7200 e Cisco serie 2600/3600, supportano WFQ (Weighted Fair Queueing) nella classe predefinita; tutte le classi di larghezza di banda utilizzano FIFO (First In First Out).

D. Quali comandi è possibile utilizzare per monitorare le code?

R. Utilizzare i seguenti comandi per monitorare le code:

- show queue {interface}{interface number} - Su piattaforme Cisco IOS diverse da Cisco serie 7500, questo comando visualizza le code o le conversazioni attive. Se l'interfaccia o il circuito virtuale (VC) non è congestionato, non verrà elencata alcuna coda. Sul Cisco serie 7500, il comando show queue non è supportato.
- show queueing interface-number [vc [[vpi/] vci] - Visualizza le statistiche di coda di

un'interfaccia o di un VC. Anche quando non ci sono congestioni, si sarà comunque in grado di vedere alcuni colpi qui. Il motivo è che i pacchetti con commutazione di contesto vengono sempre conteggiati indipendentemente dalla congestione che si verifica. Il conteggio dei pacchetti Cisco Express Forwarding (CEF) e Fast-Switched non viene effettuato a meno che non ci sia congestione. I meccanismi di coda legacy come Priority Queueing (PQ), Custom Queueing (CQ) e Weighted Fair Queueing (WFQ) non forniscono statistiche di classificazione. Nelle immagini successive alla versione 12.0(5)T, solo le funzionalità modulari basate su MQC (Quality of Service Command Line Interface) forniscono queste statistiche.

- show policy interface {interface}{interface number} - Il contatore `packets` conta il numero di pacchetti che soddisfano i criteri della classe. Questo contatore consente di aumentare la congestione dell'interfaccia. Il contatore `Pacchetti corrispondenti` indica il numero di pacchetti corrispondenti ai criteri della classe quando l'interfaccia era congestionata. Per ulteriori informazioni sui contatori dei pacchetti, consultare il seguente documento:

  Informazioni sui contatori di pacchetti nell'output dell'interfaccia della mappa dei criteri

- Cisco Class-Based QoS Configuration and Statistics MIB: fornisce funzionalità di monitoraggio SNMP (Simple Network Management Protocol).

D. RSVP può essere utilizzato in combinazione con CBWFQ (Weighted Fair Queueing) basato su classi. Quando il protocollo RSVP (Resource Reservation Protocol) e il protocollo CBWFQ sono entrambi configurati per un'interfaccia, RSVP e CBWFQ agiscono in modo indipendente, mostrando lo stesso comportamento che assumerebbero se fossero eseguiti singolarmente? RSVP sembra comportarsi come se CBWFQ non fosse configurato per quanto riguarda la disponibilità, la valutazione e l'allocazione della larghezza di banda.

A. Quando si usano RSVP e CB-WFQ nel software Cisco IOS versione 12.1(5)T e successive, il router può funzionare in modo che i flussi RSVP e le classi CBWFQ condividano la larghezza di banda disponibile su un'interfaccia o sul PVC, senza sovrascrivere.

Il software IOS versione 12.2(1)T e successive consente a RSVP di eseguire il controllo dell'ammissione utilizzando il proprio pool "ip rsvp bandwidth", mentre CBWFQ esegue la classificazione, l'applicazione di policy e la pianificazione dei pacchetti RSVP. In questo caso, si presume che i pacchetti pre-contrassegnati dal mittente e che i pacchetti non RSVP siano contrassegnati in modo diverso.

WRED (Congestion Avoidance Weighted Random Early Detection)

D. È possibile abilitare contemporaneamente WRED (Weighted Random Early Detection) e LLQ (Low Latency Queueing) o CBWFQ (Class Based Weighted Fair Queueing)?

R. Sì. L'accodamento definisce l'ordine dei pacchetti in uscita da una coda. Ciò significa che definisce un meccanismo di pianificazione dei pacchetti. Può essere utilizzato anche per garantire un'equa assegnazione della larghezza di banda e una larghezza di banda minima. La RFC (Request for Comments) 2475 definisce invece l'eliminazione come il "processo di eliminazione dei pacchetti in base a regole specificate". Il meccanismo predefinito di rilascio è tail drop, in cui l'interfaccia scarta i pacchetti quando la coda è piena. Un meccanismo alternativo di rilascio dei pacchetti è il Random Early Detection (RED) e il Cisco WRED, che iniziano il rilascio dei pacchetti in modo casuale prima che la coda sia piena e cercano di mantenere una profondità media costante. WRED utilizza il valore di precedenza IP dei pacchetti per prendere una decisione di rifiuto differenziata. Per ulteriori informazioni, fare riferimento a WRED (Weighted Random Early Detection).

D. Come è possibile monitorare WRED (Weighted Random Early Detection) e verificarne l'effettiva efficacia?

R. WRED controlla la profondità media della coda e inizia a rilasciare pacchetti quando il valore calcolato supera il valore di soglia minimo. Eseguire il comando show policy-map interface e monitorare il valore medio della profondità della coda, come mostrato nell'esempio seguente:

<#root>

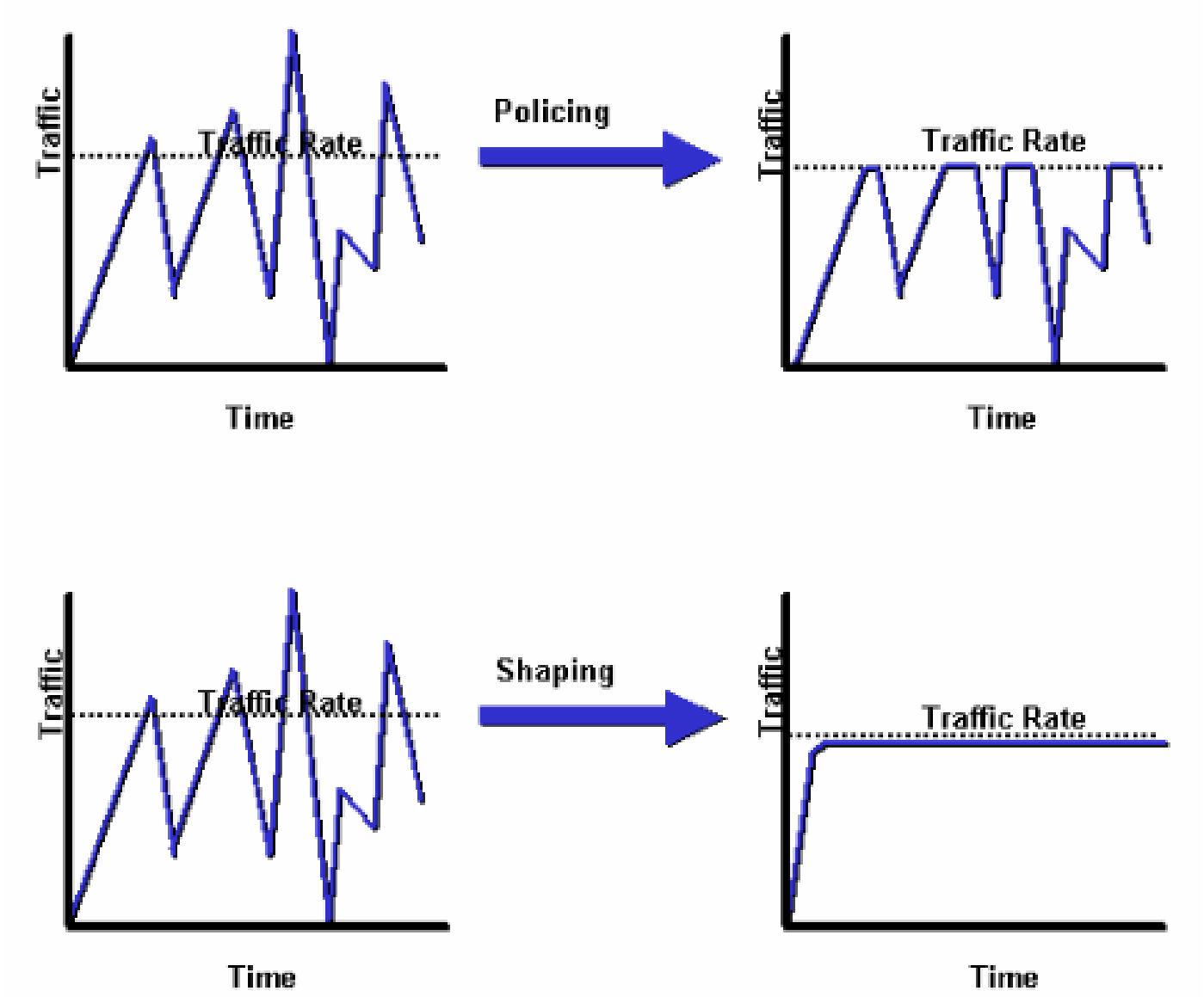
```
Router#
show policy interface s2/1


  Serial2/1
  output : p1
   Class c1
    Weighted Fair Queueing
        Output Queue: Conversation 265
          Bandwidth 20 (%)
          (pkts matched/bytes matched) 168174/41370804
          (pkts discards/bytes discards/tail drops) 20438/5027748/0
          mean queue depth: 39

     Dscp      Random drop       Tail drop        Minimum   Maximum     Mark
     (Prec)     pkts/bytes        pkts/bytes      threshold threshold probability
       0(0)     2362/581052        1996/491016        20        40       1/10
       1           0/0                0/0             22        40       1/10
       2           0/0                0/0             24        40       1/10
       [output omitted]
```

Policing and Shaping

D. Qual è la differenza tra sorveglianza e formazione?

R. Il diagramma seguente illustra la differenza fondamentale. Il Traffic Shaping conserva i pacchetti in eccesso in una coda e quindi li pianifica per trasmissioni successive con incrementi di

tempo. Il risultato del traffic shaping è una velocità di trasmissione dei pacchetti fluida e uniforme. Al contrario, il traffic policing propaga delle esplosioni. Quando la velocità del traffico raggiunge la velocità massima configurata, il traffico in eccesso viene eliminato (o contrassegnato diversamente). Il risultato è una velocità di trasmissione che appare come un'onda a dente di sega con picchi ad andamento positivo e negativo.



Per ulteriori informazioni, consultare Cenni preliminari su Policing and Shaping.

D. Che cos'è un token bucket e come funziona l'algoritmo?

R. Un token bucket non dispone di criteri di eliminazione o priorità. Di seguito è riportato un esempio di come funziona la metafora del token bucket:

- I token vengono inseriti nel bucket a una determinata velocità.
- Ogni token rappresenta l'autorizzazione per l'origine a inviare un determinato numero di bit.
- Per inviare un pacchetto, il regolatore del traffico deve essere in grado di rimuovere dal bucket un certo numero di token uguale alla dimensione del pacchetto in termini di

rappresentazione.

- Se nel bucket non è disponibile un numero di token sufficiente per inviare un pacchetto, quest'ultimo attende fino a quando il bucket non ha abbastanza token (nel caso del traffic shaping) o il pacchetto non viene ignorato o contrassegnato (nel caso del traffic policing).
- Il bucket stesso ha una capacità specificata. I token che arrivano dopo che il bucket ha raggiunto la sua capacità massima vengono ignorati e non risultano disponibili per i pacchetti futuri. Pertanto, in qualsiasi momento, il burst di dimensioni maggiori che una sorgente può inviare alla rete è all'incirca proporzionale alla dimensione del bucket. Un token bucket consente la burstiness, ovvero l'aumento e la diminuzione intermittenti della frequenza, ma la limita.

D. Con un'autorità di controllo del traffico come la policy basata su classi, cosa significano Committed Burst (BC) ed Excess Burst (Be) e come è possibile selezionare questi valori?

R. Un controllore del traffico non deposita i pacchetti in eccesso e li trasmette successivamente, come nel caso di uno shaper. Al contrario, il policer esegue un'operazione di invio semplice o non invia criteri senza memorizzazione nel buffer. Durante i periodi di congestione, poiché non è possibile memorizzare nel buffer, la soluzione migliore è scartare i pacchetti in modo meno aggressivo configurando correttamente la frammentazione estesa. Pertanto, è importante comprendere che il policer utilizza i valori di burst normale e di burst esteso per garantire il raggiungimento del CIR (Committed Information Rate) configurato.

I parametri di frammentazione sono modellati in modo approssimativo sulla regola di buffer generica per i router. La regola consiglia di configurare il buffering in modo che equivalga al bit rate del tempo di andata e ritorno per adattare le finestre TCP (Transmission Control Protocol) in sospeso di tutte le connessioni in periodi di congestione.

La tabella riportata di seguito descrive lo scopo e la formula consigliata per i valori di frammentazione normale ed estesa.

Parametro burst	Scopo	Formula consigliata
burst normale	• Implementa un bucket di token standard. • Imposta le dimensioni massime del bucket di token (sebbene i token possano essere	<#root> **CIR [BPS] * (1 byte)/(8 bits) * 1.5 seconds** Nota: 1,5 secondi è il tempo di andata e ritorno tipico.

	presi in prestito se Be è maggiore di BC). • Determina le dimensioni massime del bucket di token, in quanto i token appena in arrivo vengono eliminati e non sono disponibili per i pacchetti futuri se il bucket esaurisce la capacità.	
burst esteso	• Implementa un token bucket con funzionalità burst estesa. • Disattivato impostando BC = Be. • Quando BC è uguale a Be, il regolatore del traffico non può prendere in prestito i token e semplicemente scarta il pacchetto quando non sono disponibili token sufficienti.	<#root> **2 * normal burst**

Non tutte le piattaforme utilizzano o supportano lo stesso intervallo di valori per un policer. Per informazioni sui valori supportati per la piattaforma in uso, consultare il documento seguente:

- Panoramica di Policing and Shaping

D. In che modo CAR (Committed Access Rate) o il policing basato su classi decide se un pacchetto è conforme o supera il valore CIR (Committed Information Rate)? Il router rifiuta i pacchetti e segnala una frequenza di superamento anche se la frequenza conformata è inferiore al CIR configurato.

R. Un regolatore del traffico utilizza i valori di burst normale e di burst esteso per garantire il raggiungimento del CIR configurato. Impostare valori di burst sufficientemente elevati è importante per garantire un buon throughput. Se i valori di burst sono configurati in modo troppo basso, la velocità raggiunta può essere molto inferiore alla velocità configurata. La punizione di burst temporanei può avere un forte impatto negativo sulla velocità di trasmissione del traffico TCP (Transmission Control Protocol). Con CAR, eseguire il comando show interface rate-limit per monitorare il burst corrente e determinare se il valore visualizzato è sempre vicino ai valori di limite (BC) e limite esteso (Be).

<#root>

```
rate-limit 256000 7500 7500 conform-action continue exceed-action drop
rate-limit 512000 7500 7500 conform-action continue exceed-action drop


router#

show interfaces virtual-access 26 rate-limit

    Virtual-Access26 Cable Customers
      Input
        matches: all traffic
          params:  256000 BPS, 7500 limit, 7500 extended limit
          conformed 2248 packets, 257557 bytes; action: continue
          exceeded 35 packets, 22392 bytes; action: drop
          last packet: 156ms ago, current burst: 0 bytes
          last cleared 00:02:49 ago, conformed 12000 BPS, exceeded 1000 BPS
      Output
        matches: all traffic


params:  512000 BPS

, 7500 limit, 7500 extended limit
          conformed 3338 packets, 4115194 bytes; action: continue
          exceeded 565 packets, 797648 bytes; action: drop
          last packet: 188ms ago,

current burst: 7392 bytes


          last cleared 00:02:49 ago,

conformed 194000 BPS, exceeded 37000 BPS
```

Per ulteriori informazioni, consultare i seguenti documenti:

- Panoramica di Policing and Shaping

- Sorveglianza QoS su Catalyst 6000
- Domande frequenti sulla Quality of Service su Catalyst 4000
- Domande frequenti sugli switch Catalyst serie G-L3 e sui moduli QoS WS-X4232-L3 Layer 3

D. La frammentazione e il limite della coda sono indipendenti l'uno dall'altro?

R. Sì, i burst del policer e il limite della coda sono separati e indipendenti l'uno dall'altro. È possibile visualizzare il policer come un gate che consente un certo numero di pacchetti (o byte) e la coda come un bucket di dimensioni limite della coda che contiene i pacchetti ammessi prima della trasmissione sulla rete. In teoria, si desidera che il bucket sia abbastanza grande da contenere una raffica di byte/pacchetti ammessi dal gate (policer).

Frame Relay QoS (Quality of Service)

D. Quali valori è necessario selezionare per le opzioni CIR (Committed Information Rate), BC (Committed Burst), Be (Excess Burst) e MinCIR (Minimum CIR)?

R. Frame Relay Traffic Shaping, abilitato usando il comando frame-relay traffic-shaping, supporta diversi parametri configurabili. Questi parametri includono `frame-relay cir`, `frame-relay mincir` e `frame-relay BC`. Per ulteriori informazioni sulla selezione di questi valori e sui comandi show correlati, consultare i seguenti documenti:

- Configurazione di Frame Relay Traffic Shaping su router 7200 e piattaforme inferiori
- Mostra comandi per Frame Relay Traffic Shaping
- VoIP su Frame Relay con qualità del servizio (frammentazione, Traffic Shaping, priorità IP RTP)

D. La funzione Priority Queueing sull'interfaccia principale Frame Relay funziona in Cisco IOS 12.1?

A. Le interfacce Frame Relay supportano sia i meccanismi di coda delle interfacce che i meccanismi di coda per circuito virtuale (VC). A partire dalla versione Cisco IOS 12.0(4)T, la coda di interfaccia supporta FIFO (First In First Out) o PIPQ (Per Interface Priority Queueing) solo quando si configura Frame Relay Traffic Shaping (FRTS). Pertanto, la seguente configurazione non funzionerà più se si esegue l'aggiornamento a Cisco IOS 12.1.

```
interface Serial0/0
  frame-relay traffic-shaping
  bandwidth 256
  no ip address
  encapsulation frame-relay IETF
  priority-group 1

 !
```

```
 interface Serial0/0.1 point-to-point
  bandwidth 128
  ip address 136.238.91.214 255.255.255.252
  no ip mroute-cache
  traffic-shape rate 128000 7936 7936 1000
  traffic-shape adaptive 32000
  frame-relay interface-dlci 200 IETF
```

Se FRTS non è abilitato, è possibile applicare un metodo di coda alternativo, ad esempio CBWFQ (Class Based Weighted Fair Queueing), all'interfaccia principale che agisce come una pipe a larghezza di banda singola. Inoltre, a partire dalla versione Cisco IOS 12.1.1(T), è possibile abilitare Frame Relay Permanent Virtual Circuits (PVC) Priority Interface Queueing (PIPQ) su un'interfaccia principale Frame Relay. È possibile definire PVC di alta, media, normale o bassa priorità ed eseguire il comando frame-relay interface-queue priority sull'interfaccia principale, come mostrato nell'esempio che segue:

<#root>

```
interface Serial3/0
 description framerelay main interface
 no ip address
 encapsulation frame-relay
 no ip mroute-cache
 frame-relay traffic-shaping


frame-relay interface-queue priority


interface Serial3/0.103 point-to-point
 description frame-relay subinterface
 ip address 1.1.1.1 255.255.255.252
 frame-relay interface-dlci 103
  class frameclass

map-class frame-relay frameclass
 frame-relay adaptive-shaping becn
 frame-relay cir 60800
 frame-relay BC 7600
 frame-relay be 22800
 frame-relay mincir 8000
 service-policy output queueingpolicy


frame-relay interface-queue priority low
```

D. Frame Relay Traffic Shaping (FRTS) funziona con Distributed Cisco Express Forwarding (dCEF) e Distributed Class Based Weighted Fair Queueing (dCBWFQ)?

R. A partire dalla versione Cisco IOS 12.1(5)T, solo la versione distribuita delle funzionalità QoS è supportata sui VIP nella serie Cisco 7500. Per abilitare il traffic shaping sulle interfacce Frame Relay, utilizzare DTS (Distributed Traffic Shaping). Per ulteriori informazioni, consultare i seguenti

documenti:

- Versatile Interface Processor-Based Distributed FRF.11 e FRF.12 per Cisco IOS release 12.1 T
- Cisco serie 7500 Frame Relay Traffic Shaping con QoS distribuito

Quality of Service (QoS) su ATM (Asynchronous Transfer Mode)

D. Dove è possibile applicare una policy sui servizi con CBWFQ (Weighted Fair Queueing) basato su classi e LLQ (Low Latency Queueing) su un'interfaccia ATM (Asynchronous Transfer Mode)?

R. A partire dalla versione Cisco IOS 12.2, le interfacce ATM supportano le policy sui servizi a tre livelli o interfacce logiche: interfaccia principale, sottointerfaccia e PVC (Permanent Virtual Circuit). L'applicazione del criterio dipende dalla funzionalità QoS (Quality of Service) che si sta abilitando. Le policy di coda devono essere applicate ai singoli circuiti virtuali (VC) poiché l'interfaccia ATM monitora il livello di congestione per ogni VC e mantiene le code per i pacchetti in eccesso per ogni VC. Per ulteriori informazioni, consultare i seguenti documenti:

- Dove è possibile applicare una policy sui servizi QoS su un'interfaccia ATM?
- Informazioni su Accodamento trasmissione per VC su interfacce PA-A3 e NM-1A ATM

D. Quali byte vengono conteggiati dalle code IP to Asynchronous Transfer Mode (ATM) Class of Service (CO)?

A. I comandi di larghezza di banda e priorità configurati in una policy del servizio per abilitare, rispettivamente, CBWFQ (Class-Based Weighted Fair Queueing) e LLQ (Low Latency Queueing), utilizzano un valore Kbps che conteggia gli stessi byte di sovraccarico conteggiati dall'output del comando show interface. In particolare, il sistema di coda di layer 3 conta Logical Link Control / Subnetwork Access Protocol (LLC/SNAP). Non conta quanto segue:

- Rimorchio ATM Adaptation Layer 5 (AAL5)
- Spaziatura interna per rendere l'ultima cella un multiplo pari di 48 byte
- Intestazione di cella a cinque byte
- Quali byte vengono conteggiati dall'IP per l'accodamento dei CO ATM?

D. Quanti circuiti virtuali (VCS) possono supportare una policy sui servizi contemporaneamente?

R. Il documento seguente fornisce utili linee guida sul numero di VCS ATM (Asynchronous Transfer Mode) che possono essere supportati. Sono stati implementati da 200 a 300 PVC (Permanent Virtual Circuit) VBR-nrt in modo sicuro:

- Guida alla progettazione di classi di servizio IP - ATM

Considerare inoltre quanto segue:

- Utilizza un processore potente. Ad esempio, un VIP4-80 offre prestazioni notevolmente superiori rispetto a un VIP2-50.
- Quantità di memoria pacchetto disponibile. Su NPE-400, fino a 32 MB (in un sistema con 256 MB) sono riservati al buffer dei pacchetti. Per un NPE-200, su un sistema con 128 MB vengono riservati fino a 16 MB di memoria buffer.
- Sono state sottoposte a test approfonditi le configurazioni con WRED (Weighted Random Early Detection) per-VC che funzionano simultaneamente su un massimo di 200 PVC ATM. La quantità di memoria del pacchetto nell'indirizzo VIP2-50 che può essere utilizzata per le code per-VC è limitata. Ad esempio, un VIP2-50 con 8 MB di SRAM fornisce buffer di pacchetto 1085 disponibili per la coda di CO IP-ATM per VC su cui opera WRED. Se fossero stati configurati 100 PVC ATM e se tutti i PVC dovessero sperimentare contemporaneamente un'eccessiva congestione (come potrebbe essere simulato in ambienti di test in cui verrebbe utilizzata un'origine non controllata dal flusso TCP), in media ogni PVC avrebbe circa 10 pacchetti di buffer, che potrebbero essere troppo corti per il corretto funzionamento di WRED. I dispositivi VIP2-50 con SRAM di grandi dimensioni sono quindi consigliati in progetti con un elevato numero di PVC ATM in esecuzione per ogni VC WRED e che possono sperimentare la congestione simultanea.
- Più alto è il numero di PVC attivi configurati, minore sarà il valore SCR (Sustained Cell Rate) e, di conseguenza, più breve sarà la coda richiesta da WRED per funzionare sul PVC. Pertanto, come quando si utilizzano i profili WRED predefiniti della funzionalità IP to ATM Class of Service (CO) fase 1, la configurazione di soglie di perdita WRED inferiori quando si attivano i WRED per VC su un numero molto elevato di PVC ATM congestionati a bassa velocità ridurrebbe il rischio di carenza di buffer sul VIP. La carenza di buffer nell'indirizzo VIP non causa problemi di funzionamento. In caso di carenza di buffer nel VIP, la funzione IP to ATM CO Phase 1 degrada semplicemente in First In First Out (FIFO) tail drop durante il periodo di carenza di buffer (ossia, la stessa policy di drop che si verificherebbe se la funzione IP to ATM COs non fosse attivata su questo PVC).
- Numero massimo di VCS simultanei ragionevolmente supportati.

D. Quale hardware ATM (Asynchronous Transfer Mode) supporta le funzionalità IP to ATM Class of Service (CO), tra cui CBWFQ (Class Based Weighted Fair Queueing) e LLQ (Low Latency Queueing)?

R. I codici IP to ATM si riferiscono a una serie di funzionalità abilitate su base per circuito virtuale (VC). Data questa definizione, i CO IP-ATM non sono supportati sui processori di interfaccia ATM (AIP), PA-A1 o 4500 ATM. Questo hardware ATM non supporta le code per-VC in quanto definite dall'PA-A3 e dalla maggior parte dei moduli di rete (diversi dall'ATM-25). Per ulteriori informazioni, consultare il documento seguente:

- Informazioni sul supporto hardware ATM per IP su CO ATM
- Accodamento equo ponderato basato su classi per-VC su piattaforme basate su RSP
- CBWFQ (Weighted Fair Queuing) per VC basato su classi di VC su router Cisco 7200, 3600 e 2600
- Accodamento per-VC su PA-A3-8T1/E1 IMA ATM Port Adapter
- Configurazione della coda ATM per VC su MC3810

QoS (Voice and Quality of Service)

D. Come funziona la frammentazione e l'interfoliazione del collegamento (LFI)?

R. Il traffico interattivo, ad esempio Telnet e Voice over IP, è soggetto a una maggiore latenza quando la rete elabora pacchetti di grandi dimensioni, ad esempio il protocollo FTP (File Transfer Protocol), su una rete WAN. Il ritardo dei pacchetti per il traffico interattivo è significativo quando i pacchetti FTP vengono messi in coda su collegamenti WAN più lenti. È stato concepito un metodo per frammentare pacchetti più grandi e mettere in coda i pacchetti (voce) più piccoli tra i frammenti dei pacchetti più grandi (FTP). I router Cisco IOS supportano diversi meccanismi di frammentazione di livello 2. Per ulteriori informazioni, consultare i seguenti documenti:

- Panoramica sui meccanismi di efficienza dei collegamenti
- VoIP su Frame Relay con qualità del servizio (frammentazione, Traffic Shaping, priorità IP RTP)
- Collegamenti VoIP over PPP con Quality of Service (priorità LLQ / IP RTP, LFI, cRTP)

D. Quali strumenti posso utilizzare per monitorare le prestazioni Voice over IP?

R. Cisco offre attualmente diverse opzioni per il monitoraggio della qualità del servizio (QoS) nelle reti che utilizzano soluzioni Voice over IP di Cisco. Queste soluzioni non misurano la qualità della voce utilizzando il PSQM (Perceptual Speech Quality Measurement) o alcuni dei nuovi algoritmi proposti per la misurazione della qualità della voce. A tale scopo sono disponibili gli strumenti di Agilent (HP) e NetIQ. Tuttavia, Cisco offre strumenti che forniscono un'idea della qualità della voce che si sta sperimentando misurando il ritardo, l'jitter e la perdita di pacchetti. Per ulteriori informazioni, fare riferimento a Utilizzo di Cisco Service Assurance Agent e Internetwork Performance Monitor per gestire la qualità del servizio nelle reti Voice over IP.

D. %SW_MGR-3-CM_ERROR_FEATURE_CLASS: Errore della funzionalità di Connection Manager: SSS classe: (QoS) - errore di installazione, ignorare.

R. L'errore di installazione della funzionalità osservato è un comportamento previsto quando a un modello viene applicata una configurazione non valida. Indica che i criteri del servizio non sono stati applicati a causa di un conflitto. In generale, non è consigliabile configurare il shaping in base

al valore predefinito della classe del criterio figlio in mappe di criteri gerarchiche, bensì configurarlo in base al criterio padre dell'interfaccia. Di conseguenza, questo messaggio viene stampato insieme al comando traceback.

Con le policy basate su sessione, il shaping su class-default deve essere eseguito solo a livello di sottointerfaccia o PVC. Shaping sull'interfaccia fisica non supportato. Se la configurazione viene eseguita sull'interfaccia fisica, il verificarsi di questo messaggio di errore è previsto.

Nel caso di LNS, un altro motivo potrebbe essere che i criteri del servizio potrebbero essere forniti tramite il server RADIUS quando le sessioni vengono avviate. Usare il comando show tech per visualizzare la configurazione del server radius e tutti i criteri di servizio non validi installati tramite il server radius all'accensione della sessione o negli flap.

Informazioni correlate

- Nozioni di base sull'ottimizzazione delle prestazioni
- Supporto Quality of Service (QoS)
- Documentazione e supporto tecnico – Cisco Systems

Informazioni su questa traduzione

Cisco ha tradotto questo documento utilizzando una combinazione di tecnologie automatiche e umane per offrire ai nostri utenti in tutto il mondo contenuti di supporto nella propria lingua. Si noti che anche la migliore traduzione automatica non sarà mai accurata come quella fornita da un traduttore professionista. Cisco Systems, Inc. non si assume alcuna responsabilità per l’accuratezza di queste traduzioni e consiglia di consultare sempre il documento originale in inglese (disponibile al link fornito).