Perguntas mais freqüentes sobre QoS

Contents

Introduction

Este documento abrange as Perguntas Mais Frequentes (FAQs) relacionados à Qualidade de Serviço (QoS).

General

P. O que é Qualidade de Serviço (QoS)?

A. QoS se refere à capacidade de uma rede de fornecer um serviço melhor para o tráfego de rede selecionado em várias tecnologias subjacentes, incluindo Frame Relay, modo de transferência assíncrona (ATM), redes Ethernet e 802.1, SONET e redes IP-routed.

QoS é uma coleção de tecnologias que permitem que aplicativos requisitem e recebam níveis de serviços previsíveis em termos de capacidade de throughput de dados (largura de banda), variações de latência (jitter) e retardo. Em especial, os recursos QoS fornecem um serviço de rede melhor e mais previsível através dos seguintes métodos:

- Suporte à largura de banda dedicada.
- Melhoria das características de perda.
- Impedindo e gerenciando o congestionamento de rede.
- Modelagem do tráfego de rede.
- Definindo prioridades de tráfego na rede.

O Internet Engineering Task Force (IETF) define as duas arquiteturas a seguir para o QoS:

- Serviços Integrados (IntServ)
- Serviços diferenciados (DiffServ)

O IntServ usa o protocolo de reserva de recursos (RSVP, Resource Reservation Protocol) para

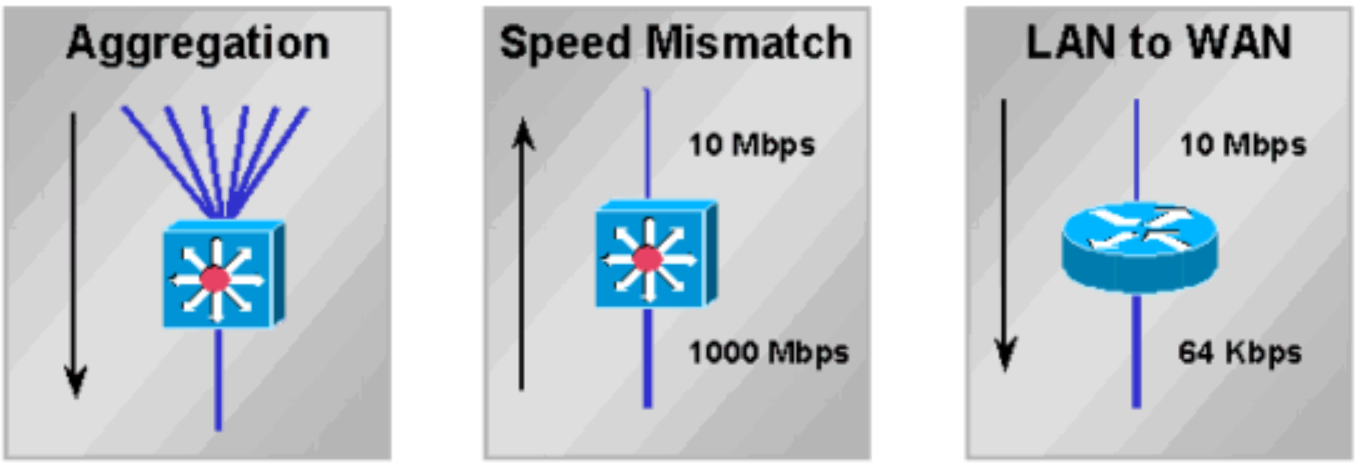
sinalizar explicitamente as necessidades de QoS de tráfego de um aplicativo com os dispositivos no caminho de ponta a ponta através da rede. Se cada dispositivo da rede ao longo do caminho puder reservar a largura de banda necessária, o aplicativo de origem pode iniciar a transmissão. A Solicitação de comentários (RFC) 2205 define RSVP, e RFC 1633 define IntServ.

O DiffServ se concentra na QoS agregada e provisionada. Em vez de sinalizar as exigências de QoS do aplicativo, o DiffServ usa um Ponto de Código DiffServ (DSCP) no cabeçalho de IP para indicar os níveis de QoS necessários. O Cisco IOS® Software Release 12.1(5)T introduziu a conformidade DiffServ nos Cisco routers. Para obter mais informações, consulte os seguintes documentos:

- Serviço integrado no Cisco IOS 12.1
- Implementing DiffServ for End-to-End Quality of Service (Implementando o DiffServ para a qualidade de serviço de ponta a ponta)
- Implementando políticas de qualidade do serviço com DSCP

P. O que são congestionamento, retardo e variação de sinal?

A. Uma interface experimenta um congestionamento quando lhe é apresentado mais tráfego do que ela consegue lidar. Os pontos de congestão da rede são fortes candidatos para mecanismos de Qualidade de Serviço (QoS). A seguir está um exemplo de pontos de congestão típicos:



O congestionamento de rede resulta em atraso. Uma rede e seus dispositivos apresentam vários tipos de atrasos, conforme explicado em Noções básicas sobre o atraso em redes de voz de pacote. A variação no retardo é conhecida como jitter, conforme explicado em Entendendo jitter nas redes de voz de pacote de informação (plataformas Cisco IOS). O atraso e o jitter precisam ser gerenciados e minimizados para suportar o tráfego em tempo real e interativo.

P. O que é MQC?

A. MQC representa a interface de linha de comando (CLI) da qualidade de serviço (QoS) modular. Ele foi criado para simplificar a configuração da QoS nos roteadores e switches Cisco, definindo uma sintaxe de comando comum e resultando em um conjunto de comportamentos de QoS entre plataformas. Este modelo substitui o modelo anterior de definição de sintaxes exclusivas para cada recurso QoS e para cada plataforma.

O MQC contém os três seguintes passos:

1. Defina uma classe de tráfego executando o comando **class-map.**
2. Crie uma política de tráfego, associando a classe de tráfego a um ou mais recursos de QoS por meio da emissão do comando policy-map.
3. Anexe a política de tráfego à interface, subinterface ou Circuito Virtual (CV), executando o comando **service-policy.**

Observação: você implementa as funções de condicionamento de tráfego do DiffServ, como marcação e modelagem, usando a sintaxe MQC.

Para obter mais informações, consulte Interface de linha de comando da qualidade de serviço modular.

P. O que a mensagem a política de serviço é suportada somente em interfaces VIP com DCEF habilitado significa?

A. Em Processadores de interface versátil (VIPs) em um Cisco 7500 Series, somente os recursos distribuídos da qualidade de serviço (QoS) são compatíveis a partir do Cisco IOS 12.1(5)T, 12.1(5)E e 12.0(14)S. A ativação do Cisco Express Forwarding distribuído (dCEF) ativa automaticamente a QoS distribuída.

Interfaces não-VIP, conhecidas como IPs (Processadores de interface) herdados, suportam recursos de QoS central como ativados no RSP (Processador de rota/Switch). Para obter mais informações, consulte os seguintes documentos:

- Enfileiramento justo ponderado distribuído de acordo com a classe e Detecção antecipada aleatória distribuída ponderada
- Enfileiramento de baixa latência distribuído
- Molde de tráfego distribuído
- FRF.11 e FRF.12 distribuídos com base no processador de interface versátil para o Cisco IOS versão 12.1 T

P. Uma política da qualidade de serviço (QoS) é compatível com quantas classes?

A. Em versões do Cisco IOS anteriores a 1.2.2, é possível definir um máximo de 256 classes e seriam definidas até 256 classes dentro de cada política se as mesmas classes forem reutilizadas para políticas diferentes. Se você tiver duas políticas, o número total de classes de ambas as políticas não deve exceder 256. Se uma política incluir o enfileiramento justo ponderado de acordo com a classe (CBWFQ) (ou seja, contém uma instrução de largura de banda [ou prioridade] em qualquer uma das classes), o número total de classes comátíveis é 64.

No CISCO IOS versões 12.2(12), 12.2(12)T e 12.2(12)S, essa limitação de 256 mapas de classe globais foi alterada e agora é possível configurar até 1024 mapas de classe globais e usar 256 mapas de classe dentro do mesmo mapa de política.

P. Como as atualizações de roteamento e keepalives de Protocolo ponto-a-ponto (PPP) / Controle de enlace de dados de alto nível (HDLC) são processados quando uma política de serviço é aplicada?

A. Os roteadores Cisco IOS utilizam os seguintes mecanismos para priorizar pacotes de controle:

- Precedência de IP
- pak_priority

Ambos os mecanismos foram projetados para garantir que os principais pacotes de informações de controle não sejam soltos ou sejam soltos por último pelo roteador e pelo sistema de fila quando uma interface externa estiver congestionada. Para obter mais informações, consulte Entendendo como Atualizações de Roteamento e Pacotes de Controle são Enfileirados em uma Interface com uma Política de Serviço QoS.

P. A Qualidade de serviço (QoS) é compatível com as interfaces configuradas com Roteamento e ponte integrados (IRB)?

A. Não. Não é possível configurar recursos de QoS quando a interface está configurada para IRB.

Classificação e marcação

P. O que é pré-classificação da Qualidade de Serviço (QoS)?

A. A pré-classificação de QoS permite corresponder e classificar o conteúdo do cabeçalho IP original de pacotes submetidos ao encapsulamento de túnel e/ou criptografia. Esse recurso não descreve o processo de copiar o valor original do byte do Tipo de serviço (ToS) do cabeçalho do pacote original para o cabeçalho do túnel. Para obter mais informações, consulte os seguintes documentos:

- Configurando QoS para redes privadas virtuais
- Qualidade de Serviço para Redes Particulares Virtuais, Módulo de Recurso 12.2(2)

P. Quais campos de cabeçalho de pacotes podem ser marcados? Quais valores estão disponíveis?

A. O recurso de marcação com base em classe permite que você defina ou marque o cabeçalho da MPLS (switching de rótulo multiprotocolo), da camada 2 ou camada 3 dos seus pacotes. Para obter mais informações, consulte os seguintes documentos:

- Configuração de marcação de pacotes de acordo com a classe
- Quando um roteador define o bit CLP em uma célula ATM?
- Configurando a marcação de pacote em PVCs de Frame Relay

P. Posso dar prioridade ao tráfego com base no URL?

A. Yes. O NBAR (Reconhecimento de aplicativo baseado em rede) permite que você classifique pacotes, comparando campos na camada do aplicativo. Antes da introdução do NBAR, a classificação mais granular consistia nos números de porta do Protocolo TCP (Transmission Control Protocol) e Protocolo UDP (User Datagram Protocol) de camada 4. Para obter mais informações, consulte os seguintes documentos:

- Perguntas e respostas sobre reconhecimento de aplicativos com base na rede
- Rede de aplicação de NBAR
- Utilizando Reconhecimento de Aplicativos Baseados em Rede e Listas de Controle de

Acesso para Bloqueio de Worm de Código Vermelho
- Como proteger sua rede contra o vírus Nimda

P. Que plataformas e versões de Cisco IOS Software suportam reconhecimento de aplicativo baseado em rede (NBAR)?

A. O suporte para NBAR é apresentado nas seguintes versões do software Cisco IOS:

Platform	Versão mínima do Cisco IOS Software
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

Observação: você precisa habilitar o Cisco Express Forwarding (CEF) para usar o NBAR.

O NBAR distribuído (DNBAR) está disponível nas seguintes plataformas:

Platform	Versão mínima do Cisco IOS Software
7500	12.2(4)T, 12.1(6)E
FlexWAN	12.1(6)E

Observação: o NBAR não é suportado nas interfaces de VLAN do Catalyst 6000 Multilayer Switch Feature Card (MSFC), Cisco 12000 Series ou Route Switch Module (RSM) para o Catalyst 5000 Series. Se você visualizar uma plataforma específica listada acima, entre em contato com o representante técnico da Cisco.

Gerenciamento de enfileiramento e de congestionamento

P. Qual é a finalidade do enfileiramento?

A. O enfileiramento foi criado para acomodar o congestionamento temporário na interface de um dispositivo de rede, armazenando os pacotes excedentes em buffers até que a largura de banda esteja disponível. Os roteadores Cisco IOS suportam diversos métodos de enfileiramento para satisfazer as exigências variáveis de largura de banda, variação de sinal e retardo de diferentes aplicativos.

O mecanismo padrão na maioria das interfaces é First In First Out (FIFO). Alguns tipos de tráfego têm requisitos de atraso/variação de sinal mais rígidos. Assim, um dos mecanismos alternativos de enfileiramento a seguir deve ser configurado ou está habilitado por padrão:

- Weighted-Fair Queuing (WFQ)
- Enfileiramento moderado ponderado com base em classe (CBWFQ)

- LLQ (Enfileiramento de Baixa Latência), que é na verdade o CBWFQ com uma PQ (Fila de Prioridade) (conhecida como PQCBWFQ)
- Priority Queueing (PQ)
- Custom Queueing (CQ)

O enfileiramento geralmente só acontece nas interfaces externas. Um roteador enfileira pacotes que estão saindo de uma interface. Você controlar policiar o tráfego de entrada, mas normalmente não é possível enfileirar a entrada (uma exceção é o buffer no lado do recebimento em um roteador Cisco 7500 Series que usa o Cisco Express Forwarding distribuído (dCEF) para encaminhar pacotes da interface de entrada para a interface de saída; para obter mais informações, consulte Noções básicas da execução de VIP CPU em 99% e Rx-Side Buffering. Em plataformas de produto avançado distribuídas, como as séries Cisco 7500 e 12000, a interface de entrada pode utilizar seus próprios buffers de pacote para armazenar tráfego em excesso comutado para uma interface de saída congestionada seguindo a decisão de switching da interface de entrada. Em condições raras, geralmente quando a interface de entrada está alimentando uma interface de saída mais lenta, a interface de entrada pode passar por erros ignorados de acréscimo quando ela funciona fora da memória do pacote. O excesso de congestionamento pode levar a quedas de filas de saída. Quedas de fila de entrada têm uma causa-raiz diferente na maioria das vezes. Para obter mais informações sobre Troubleshooting de queda, consulte o seguinte documento:

- Troubleshooting de Quedas de Fila de Entrada e Quedas de Fila de Saída

Para obter mais informações, consulte os seguintes documentos:

- Troubleshooting de Erros "Ignorados" em um ATM Port Adapter
- Troubleshooting de Erros Ignorados e Quedas Sem Memória no Cisco 12000 Series Internet Router

P. Como os padrões WFQ e CBWFQ funcionam?

A. O enfileiramento justo busca alocar uma parte justa da largura de banda da interface entre conversas ativas ou fluxos IP. Ele classifica os pacotes em subfilas, identificados por um número de identificação de conversação, usando um algoritmo de hashing baseado em vários campos do cabeçalho de IP e no tamanho do pacote. A seguir você saberá como é calculado o peso:

- **W=K/(precedência +1)**

K= 4096 com Cisco IOS 12.0(4)T e versões anteriores, e 32384 com 12.0(5)T e versões posteriores.

Quanto menor o peso, maior é a prioridade e o compartilhamento da largura de banda. Além do peso, o comprimento do pacote é levado em consideração.

O CBWFQ permite que você defina uma classe de tráfego e atribua a ela uma garantia de largura de banda mínima. O algoritmo por trás desse mecanismo é o WFQ, o que explica o nome. Para configurar CBWFQ, defina classes específicas em instruções com classes de mapas. Então, você atribui uma política para cada classe em um mapa de políticas. Esse mapa de política será anexado externamente a uma interface. Para obter mais informações, consulte os seguintes documentos:

- Compreendendo a classe baseada em Weighted Fair Queuing em ATM
- Compreendendo a Weighted Fair Queuing em ATM

P. Se uma classe do CBWFQ (Enfileiramento Justo e Ponderado Baseado em Classes) não está usando sua largura de banda, outras classes podem usar a largura de banda?

A. Yes. Embora as garantias de largura de banda fornecidas emitindo os comandos bandwidth e priority tenham sido descritas com palavras como reserved e bandwidth to be set aside, nenhum dos comandos implementa uma reserva verdadeira. Significa que se uma classe de tráfego não estiver usando sua largura de banda configurada, qualquer largura de banda não utilizada é compartilhada entre as outras classes.

O sistema de filas impõe uma exceção importante a essa regra, com uma classe de prioridade. Conforme observado acima, a carga oferecida de uma classe de prioridade é medida por um vigilante de tráfego. Durante condições de congestionamento, uma classe de prioridade não pode usar excesso de largura de banda. Para obter mais informações, consulte Comparando os comandos bandwidth e priority de uma política de serviços de QoS.

P. O enfileiramento justo ponderado de acordo com a classe (CBWFQ) é compatível com subinterfaces?

A. As interfaces lógicas do Cisco IOS não têm suporte inerente para um estado de congestionamento e não suportam a aplicação direta de uma política de serviços que inclua um método de enfileiramento. Em vez disso, é preciso primeiro aplicar modelagem à subinterface, utilizando GTS (Modelagem de Tráfego Genérico) ou modelagem baseada em classe. Para obter mais informações, consulte Applying QoS Features to Ethernet Subinterfaces (Aplicando recursos de QoS a subinterfaces Ethernet).

P. Qual é a diferença entre as instruções de prioridade e largura de banda em um mapa de política?

A. Os comandos priority e bandwidth diferem em ambas funcionalidades e quais as aplicações eles tipicamente suportam. A tabela a seguir resume essas diferenças:

Função	Comando bandwidth	Comando priority
Garantia de largura de banda mínima	Yes	Yes
Garantia máxima de largura de banda	No	Yes
Policer embutido	No	Yes
Fornece latência baixa	No	Yes

Para obter mais informações, consulte Comparando os comandos bandwidth e priority de uma política de serviços de QoS.

P. Como é calculado o limite de fila no FlexWAN e no VIP (Processadores de Interface Versáteis)?

A. Supondo uma SRAM suficiente no VIP ou FlexWAN, o limite de fila é calculado com base em um atraso máximo de 500 ms e o tamanho de pacote médio é 250 bytes. Este é um exemplo de

uma classe com um Mbps de largura de banda:

Limite da fila = 1000000 / (250 x 8 x 2) = 250

Limites de fila menores são atribuídos à medida que a quantidade de memória de pacotes disponíveis diminui e com um número maior de Circuitos Virtuais (VCs).

No exemplo a seguir, um PA-A3 está instalado em uma placa FlexWAN para o Cisco 7600 Series e está suportando diversas subinterfaces com Circuitos virtuais permanentes (PVCs) de 2 MB. A política de serviço é aplicada a cada VC.

```
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
```

A interface do Modo de transferência assíncrono (ATM) obtém um limite de fila para toda a interface. O limite é uma função do total de buffers disponíveis, o número de interfaces físicas no FlexWAN e o retardo máximo de enfileiramento permitido na interface. Cada PVC obtém uma parte do limite da interface com base na SCR (Taxa média de células) ou na MCR (Taxa mínima de células) do PVC, e cada classe obtém uma parte do limite do PVC com base na sua alocação de largura de banda.

A seguinte saída de exemplo do comando show policy-map interface é derivada de um FlexWAN com buffers globais 3687. Emita o comando show buffer para ver esse valor. Cada PVC de 2 Mbps recebe 50 pacotes com base na largura de banda de PVC de 2 Mbps (2047/149760 x 3687 = 50). Cada classe é alocada para uma parte dos 50, como mostrado na saída a seguir:

```
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 os seus fluxos de tráfego usarem tamanhos de pacotes grandes, a saída do comando show policy-map interface poderá reportar um valor incremental para o campo no buffer drops, pois você poderá ficar sem buffers antes de atingir o limite de filas. Nesse caso, tente reduzir manualmente o limite de fila em classes não prioritárias. Para obter mais informações, consulte Noções básicas sobre a transmitir de limite de fila com IP para ATM CoS.

P. Como você verifica o valor de limite da fila?

A. Em plataformas não distribuídas, o limite de fila é 64 pacotes por padrão. A saída de exemplo a seguir foi capturada em um Cisco 3600 Series Router:

```
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
```

P. Posso habilitar o fair queueing dentro de uma classe?

A. O Cisco 7500 Series com QoS (Qualidade de serviço) distribuída suporta enfileiramento moderado por classe. Outras plataformas, incluindo Cisco 7200 Series e Cisco 2600/3600, são compatíveis com o Enfileiramento justo ponderado (WFQ) na classe class-default; todas as classes de largura de banda usam FIFO.

P. Quais comandos posso usar para monitorar o enfileiramento?

A. Use os seguintes comandos para monitorar o enfileiramento:

- **show queue {interface}{interface number}** - nas plataformas no Cisco IOS, exceto o Cisco 7500 Series, esse comando exibe a filas ou conversas ativas. Se a interface ou Circuito Virtual (VC) não estiver congestionado, nenhuma fila será listada. Na Série Cisco 7500, o comando show queue não é suportado.
- **show queueing interface interface-number [vc [[vpi/] vci]** - esse comando exibe as estatísticas de enfileiramento de uma interface ou um VC. Mesmo quando há congestionamento, ainda será possível visualizar algumas acertos aqui. A razão para isso é que os pacotes de processo comutados são sempre contados, independentemente de haver congestionamento. O Cisco Express Forwarding (CEF) e os pacotes de Fast-Switched não são contados, a menos que haja congestionamento. Os mecanismos de enfileiramento antigos, como Enfileiramento de prioridade (PQ), Enfileiramento personalizado (CQ) e Enfileiramento justo ponderado (WFQ), não fornecerão as estatísticas de classificação. Somente os recursos baseados na Interface de linha de comando da Qualidade de serviço modular (MQC) nas imagens mais recentes que 12.0(5)T fornecem essas estatísticas.
- **show policy interface {interface}{interface number} - o contador de pacotes conta o número de pacotes que correspondem aos critérios da classe.** Esse contador é incrementado estando a interface congestionada ou não. O `contador de pacotes compatíveis indica o número de pacotes que correspondem ao critério da classe, quando a interface estava congestionada.` Para obter informações adicionais sobre os contadores de pacote, consulte o seguinte documento:Compreendendo os contadores de pacotes na saída de show policy-map interface
- MIB de configuração e estatísticas da QoS de acordo com a classe da Cisco - fornece recursos de monitoramento do Protocolo de gerenciamento de rede (SNMP).

P. O RSVP pode ser usado em conjunto com o Enfileiramento justo ponderado de acordo com a classe (CBWFQ). Quando o Resource Reservation Protocol (RSVP) e o CBWFQ estão configurados para uma interface, eles atuam independentemente, exibindo o mesmo comportamento que teriam se estivessem em execução isoladamente? Parece que o RSVP se comporta como se o CBWFQ não estivesse configurado em relação à disponibilidade, avaliação e alocação da largura de banda.

A. Ao usar RSVP e CB-WFQ na versão 12.1(5)T e mais recente do Software Cisco IOS, o roteador pode operar de modo que os fluxos de RSVP e as classes de CBWFQ compartilhem a largura de banda disponível em uma interface ou PVC, sem esgotamento.

O Software IOS Versão 12.2(1)T e posterior permite que o RSVP faça o controle de admissão

usando o seu próprio pool "ip rsvp bandwidth", enquanto o CBWFQ faz a classificação, a vigilância e a programação de pacotes RSVP. Isso pressupõe pacotes pré-marcados pelo remetente e que os pacotes não-RSVP são marcados de forma diferente.

Weighted Random Early Detection (WRED) da prevenção de congestionamentos

P. Posso ativar ao mesmo tempo Weighted Random Early Detection (WRED) e Low Latency Queueing (LLQ) ou Class Based Weighted Fair Queueing (CBWFQ)?

A. Yes. O enfileiramento define a ordem dos pacotes que deixam a fila. Isto significa que ele define um mecanismo de programação de pacote. Ele também pode ser usado para fornecer garantias de alocação de largura de banda justa e largura de banda mínima. Em contrate, a Solicitação de comentário (RFC) 2475 define queda como o processo de descartar pacotes com base em regras especificadas. O mecanismo de queda padrão é a queda traseira, na qual a interface derruba pacotes quando a fila está cheia. Um mecanismo de descarte alternativo é a Detecção inicial aleatória (RED) e o WRED da Cisco, que começa a descartar pacotes aleatoriamente antes que a fila esteja cheia e procura manter uma profundidade média coerente da fila. A WRED utiliza valor de precedência de IP de pacotes para tomar uma decisão de queda diferenciada. Para obter mais informações, consulte Detecção inicial aleatória ponderada (WRED).

P. Como monitorar a Detecção inicial aleatória ponderada (WRED) e vê-la realmente em vigor?

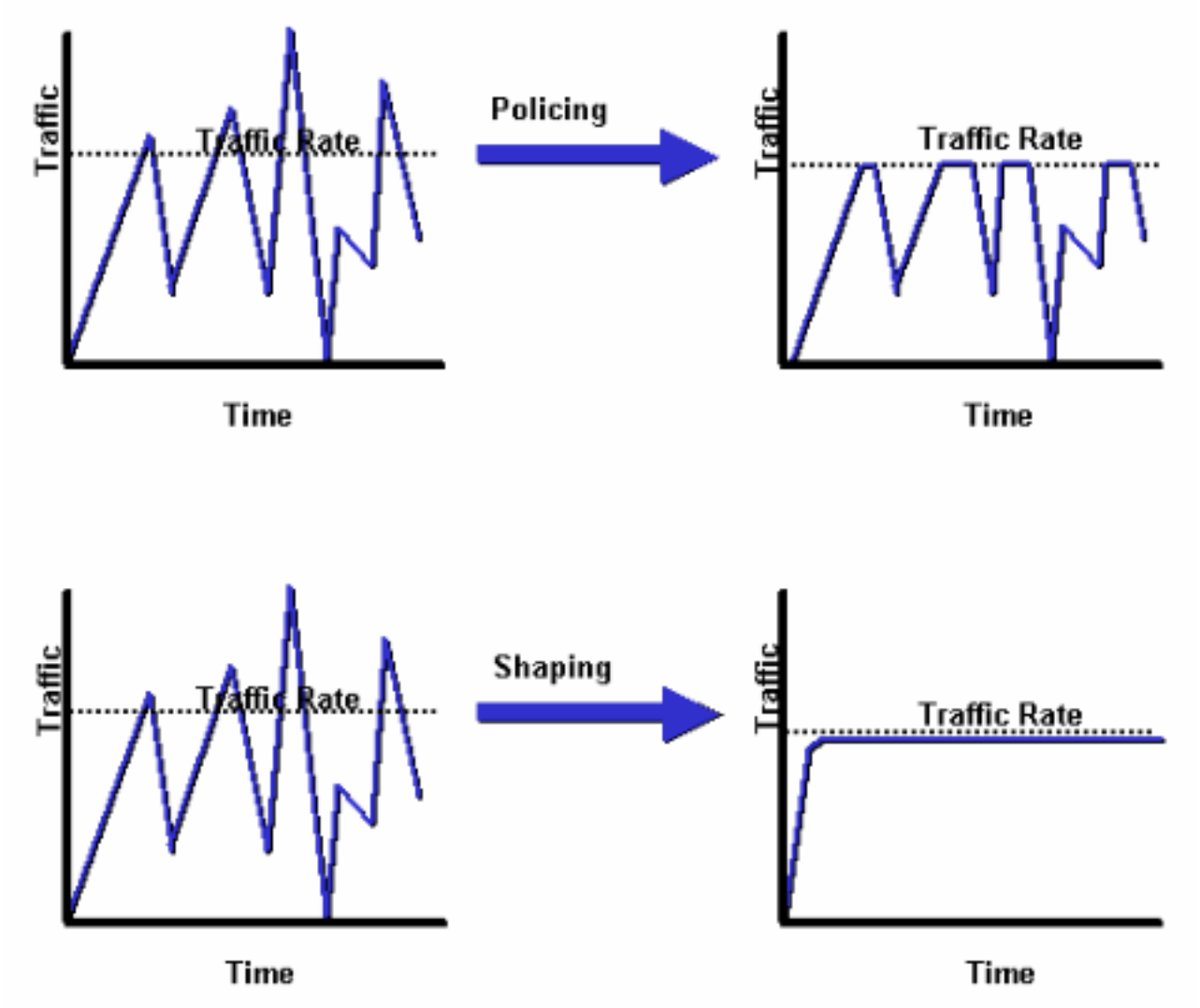
A. A WRED monitora a profundidade média da fila e começa a descartar os pacotes quando o valor calculado excede o valor de limite mínimo. Execute o comando **show policy-map interface e monitore o valor de profundidade média da fila, conforme mostrado no exemplo a seguir:**

```
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]
```

Vigilância e molde

P. Qual a diferença entre vigilância e modelagem?

A. O diagrama a seguir ilustra a diferença principal: A modelagem de tráfego retém os pacotes excedentes em uma fila e agenda o excesso para transmissão posterior em intervalos de tempo. O resultado da modelagem de tráfego é uma taxa de saída de pacote facilitada. Em contrapartida, a vigilância de tráfego propaga bursts. Quando a taxa de tráfego atinge a taxa máxima configurada, o tráfego de excesso é liberado (ou remarcado). O resultado é uma taxa de saída que aparece como um dente de serra com picos e depressões.



Para obter mais informações, consulte Resumo de policiamento e modelagem.

P. O que é um token bucket e como o algoritmo funciona?

A. Um token bucket propriamente dito não possui política de descarte ou prioridade. A seguir, há um exemplo de como funciona a metáfora do token bucket:

- Os tokens são colocados em um bucket a uma certa taxa.
- Cada token é uma permissão para que a origem envie um determinado número de bits.
- Para enviar um pacote, o regulador de tráfego deve estar apto a remover do bucket um número de símbolos que seja igual em representação ao tamanho do bucket.
- Se não houver tokens suficientes no bucket para enviar um pacote, o pacote esperará até que o bucket tenha tokens suficientes (no caso de um formador) ou o pacote será descartado ou marcado (no caso de um vigilante).
- O bucket propriamente dito possui uma capacidade específica. Se o bucket for totalmente

preenchido, os tokens recém-chegados serão descartados e não estarão disponíveis para pacotes futuros. Assim, a qualquer momento, o maior surto que uma origem pode enviar na rede é proporcional ao tamanho do pacote. Um token bucket permite intermitência, mas a limita.

P. Com um agente de tráfego, como o policiamento de acordo com a classe, o que significa Pico confirmado (BC) e Pico excedente (Be) e como devo escolher esses valores?

A. Um vigilante de tráfego não faz buffer de pacotes em excesso e transmite-os mais tarde, como é o caso de um modelador. Em vez disso, o vigilante executa um envio simples ou não envia a política sem o buffer. Durante os períodos de congestionamento, como você não pode executar o buffer, o melhor que você pode fazer é descartar pacotes de forma menos agressiva, configurando corretamente o pico prolongado. Portanto, é importante compreender que o agente usa os valores de pico normal e pico prolongado para assegurar que a Taxa de informações confirmadas (CIR) seja acessível.

Os parâmetros de intermitência são modelados livremente na regra de buffer genérico para roteadores. A regra recomenda configurar o armazenamento em buffer igual à taxa de bits do tempo de round-trip para acomodar as janelas de protocolo de controle de transmissão (TCP) pendentes de toda as conexões em períodos de congestionamento.

A tabela a seguir descreve a finalidade e a fórmula recomendada para os valores de pico normal e prolongado:

Parâmetro Burst	Propósito	Fórmula recomendada
pico normal	• Implementa um token bucket padrão. • Define o tamanho máximo do token bucket (embora tokens possam ser emprestados se Be for maior que BC). • Determina o tamanho do token bucket já que os tokens recém-chegados serão descartados e não estarão disponíveis para futuros pacotes caso o bucket seja totalmente preenchido.	`CIR [BPS] * (1 byte)/(8 bits) * 1.5 seconds` **Nota:** 1,5 segundos é o tempo de ida e volta típico.
intermitência estendida	• Implementa um token bucket com o recurso de intermitência estendida. • Desativado, definindo BC = Be. • Quando BC é igual a Be, o regulador de tráfego não pode solicitar tokens e simplesmente descarta o pacote quando não	`2 * normal burst`

	há tokens suficientes disponíveis.	

Nem todas as plataformas utilizam ou suportam o mesmo intervalo de valores para um policer. Consulte o documento a seguir para saber os valores que são suportados em sua plataforma específica:

- Visão geral de vigilância e modelagem

P. Como a Taxa de acesso confirmada (CAR) ou o policiamento de acordo com a classe decide se um pacote atende ou excede à Taxa de informações confirmada (CIR)? O roteador descarta pacotes e relata uma taxa excedida mesmo se a taxa conformada for menor que a CIR configurada.

A. Um vigilante de tráfego usa a intermitência normal e os valores de intermitência estendidos para garantir que o CIR configurado seja atingido. A definição de valores de intermitência altos o suficiente é importante para assegurar uma boa transferência. Se a intermitência for configurada com valores muito baixos, a taxa obtida poderá ser muito mais baixa do que a taxa configurada. Punir os bursts temporários pode ter um forte impacto adverso sobre o throughput do tráfego do TCP (Protocolo de controle de transmissão). Com a CAR, execute o comando **show interface rate-limit para monitorar o pico atual e determinar se o valor exibido está constantemente próximo aos valores de limite (BC) e limite estendido (Be).**

```
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
```

Para obter mais informações, consulte os seguintes documentos:

- Visão geral de vigilância e modelagem
- Vigilância de QoS no Catalyst 6000
- Perguntas freqüentes sobre qualidade de serviço no Catalyst 4000
- Perguntas frequentes sobre Catalyst G-L3 Series Switches e QoS dos módulos de camada 3 WS-X4232-L3

P. O pico e o limite de fila são independentes um do outro?

A. Sim, o pico de agente e o limite de fila são separados e independentes um do outro. Você pode ver o agente como uma entrada que permite que determinado número de pacotes (ou bytes) e a fila como um recipiente do *limite de fila do tamanho que contém os pacotes aceitos antes da transmissão na rede.* Idealmente, você deseja que seu recipiente seja grande o suficiente para armazenar um *pico de bytes/pacotes aceitos pela entrada (agente).*

Frame Relay de Qualidade de serviço (QoS)

P. Quais valores devo selecionar para Taxa de Informação Comprometida (CIR), Intermitência Comprometida (BC), Intermitência em Excesso (Be) e CIR Mínimo (MinCIR)?

A. A Modelagem de tráfego do Frame Relay, que você ativa executando o comando **frame-relay traffic-shaping, é compatível com vários parâmetros configuráveis.** Esses parâmetros incluem frame-relay cir, frame-relay mincir e frame-relay BC. Consulte os seguintes documentos para obter mais informações sobre como selecionar esses valores e compreender comandos show relacionados:

- Configuração da formatação de tráfego frame relay nos roteadores 7200 e nas plataformas inferiores
- Comandos show para a formatação de tráfego frame relay
- VoIP sobre Frame Relay com Qualidade de Serviço (fragmentação, formatação de tráfego, IP RTP Priority)

P. O enfileiramento de prioridade na interface principal do Frame Relay funciona no Cisco IOS 12.1?

A. As interfaces do Frame Relay são compatíveis com os mecanismos de enfileiramento de interface e mecanismos de enfileiramento de circuito virtual (VC). A partir do Cisco IOS 12.0(4)T, a fila de interface suporta First In First Out (FIFO) ou Per Interface Priority Queueing (PIPQ) somente quando é configurada a modelagem de tráfego de Frame Relay (FRTS). Portanto, a configuração a seguir não funcionará mais se você atualizar para o 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 o FRTS não estiver habilitado, você poderá aplicar um método padrão BWFQ, à interface

principal, que esteja atuando como uma tubulação de largura de banda única. Além disso, até o Cisco IOS 12.1.1 (T), você podia habilitar o PIPQ (Enfileiramento da Interface de Prioridade) dos PVCs (Circuitos Virtuais Permanentes de Frame Relay) em uma interface principal do Frame Relay. É possível definir PVCs com prioridade alta, média, normal ou baixa e emitir o comando frame-relay interface-queue priority na interface principal, conforme mostrado no exemplo a seguir:

```
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
```

P. Modelagem de tráfego ddo Frame Relay (FRTS) funciona com Cisco Express Forwarding distribuído (dCEF) e enfileiramento justo ponderado distribuído de acordo com a classe (dCBWFQ)?

A. A partir do Cisco IOS 12.1(5)T, somente a versão distribuída dos recursos de QoS são compatíveis com VIPs no Cisco 7500 Series. Para habilitar a formatação de tráfego nas interfaces de frame relay, use o Distributed Traffic Shaping (DTS). Para obter mais informações, consulte os seguintes documentos:

- FRF.11 e FRF.12 distribuídos com base no processador de interface versátil para o Cisco IOS versão 12.1 T
- Formatação de tráfego frame relay com QoS distribuída no Cisco 7500 Series

Qualidade de serviço (QoS) sobre ATM (modo de transferência assíncrona)

P. Onde aplico uma política de serviço com CBWFQ e LLQ em uma interface ATM?

A. A partir do Cisco IOS 12.2, as interfaces de ATM são compatíveis com políticas de serviço em três níveis ou interfaces lógicas: interface principal, subinterface e Circuito virtual permanente (PVC). O local onde você aplica a política é uma função do recurso Qualidade de Serviço (QoS) que você está habilitando. As políticas de enfileiramento devem ser aplicadas por VC (circuito

virtual), uma vez que a interface ATM monitora o nível de congestionamento por VC e mantém as filas para pacotes em excesso por VC. Para obter mais informações, consulte os seguintes documentos:

- Onde aplico uma política de serviços de QoS em uma interface ATM?
- Entendendo o enfileiramento de transmissão per-VC nas interfaces PA-A3 e NM-1A ATM

P. Quais bytes são contados pelo IP para o enfileiramento da categoria de serviço (CoS) do modo de transferência assíncrona (ATM)?

A. Os comandos de largura de banda e prioridade configurados em uma política de serviço para habilitar os recursos Class-Based Weighted Fair Queueing (CBWFQ) e Low Latency Queueing (LLQ), respectivamente, usam um valor em Kbps que contam os mesmos bytes de sobrecarga que são contados pela saída do comando show interface. Especificamente, o sistema de enfileiramento da camada 3 conta Controle lógico de enlace/Protocolo de acesso a sub-rede (LLC/SNAP). Não considera o seguinte:

- Trailer AAL5 (ATM Adaptation Layer 5)
- Preenchendo para tornar a última célula um múltiplo par de 48 bytes.
- Cabeçalho de célula de cinco bytes
- Quais bytes são contados pelo IP para enfileiramento de ATM CoS?

P. Quantos Circuitos virtuais (VCS) são compatíveis com uma política de serviço ao mesmo tempo?

A. O seguinte documento fornece orientações úteis sobre o número de VCS ATM (Modo de Transferência Assíncrono) que pode ser suportado. Aproximadamente 200 a 300 PVCs (Circuitos virtuais permanentes) VBR-nrt foram implementados com segurança:

- Guia de projeto da categoria de seviço IP para ATM

Além disso, considere o seguinte:

- Utilize um processador potente compatível. Por exemplo, um VIP4-80 oferece desempenho consideravelmente maior que um VIP2-50.
- Quantidade de memória de pacote disponível. No NPE-400, até 32 MB (em um sistema com 256 MB) são reservados para o buffer de pacotes. Para um NPE-200, são reservados até 16 MB para buffers de pacotes em um sistema com 128 MB.
- As configurações com WRED por VC operando simultaneamente em até 200 PVCs de ATM foram exaustivamente testadas. A quantidade de memória do pacote no VIP2-50 que pode ser usada para as filas por VC é finita. Por exemplo, um VIP2-50 com SRAM de 8 MB fornece 1085 buffers de pacote disponíveis para IP para enfileiramento de ATM COs por-VC no qual o WRED funciona. Se 100 PVCs ATM foram configurados e todos os VCS passaram por congestionamento excessivo ao mesmo tempo (como poderia ser simulado em ambientes de teste, nos quais a origem controlada do fluxo não-TCP estaria sendo usada), então, em média, cada PVC teria aproximadamente 10 pacotes de buffer, o que pode ser muito pouco para que o WRED funcione da maneira apropriada. Dispositivos VIP2-50 com ampla SRAM são, portanto, extremamente recomendados em designs com um grande número de PVCs ATM que execução WRED por VC e que podem passar com congestionamentos ao mesmo tempo.

- Quanto maior o número de PVCs ativos configurados, menor a Taxa de células sustentáveis (SCR) necessária e, consequentemente, menor a fila exigido pelo WRED para operar no PVC. Portanto, como ocorre ao usar perfis WRED padrão do recurso Classe de serviço (CoS) IP para ATM Fase 1, a configuração de limiares de queda WRED mais baixos, quando o WRED por VC está ativado em um número muito grande de PVCs ATM de baixa velocidade congestionados, minimizaria o risco de falta de buffer no VIP. A queda do buffer no VIP não resulta em um mau funcionamento. No caso de interrupção de buffer no VIP, o recurso de IP para ATM COS fase 1 simplesmente degrada para o descarte de final First In First Out (FIFO) durante o período de interrupção de buffer (ou seja, a mesma política de descarte que ocorreria se o recurso de COs de IP para ATM não estivesse ativado neste PVC).
- Número máximo de VCs simultâneos que podem ser suportados razoavelmente.

P. Que hardware ATM dá suporte a recursos de IP para Classe de Serviço (CoS, Class of Service) ATM como padrão CBWFQ e LLQ (Enfileiramento de Baixa Latência)?

A. COs IP para ATM se referem a um conjunto de recursos que são ativados em uma base de Circuito virtual (VC). Dada essa definição, IP para Classe de Serviço (CoS, Class of Service) ATM não tem suporte nos processadores de rede ATM 4500 e PA-A1 ou no processador AIP. Esse hardware ATM não suporta enfileiramento per-VC como o PA-A3 e a maioria dos módulos de rede (além do ATM-25) o define. Para obter mais informações, consulte o seguinte documento:

- Entendendo o suporte a hardware ATM de IP to CoS ATM
- Per-VC Class-Based, Weighted Fair Queuing em plataformas baseadas em RSP
- Weighted Fair Queuing por VC com base em classe (CBWFQ por VC) nos Cisco 7200, 3600, e 2600 Routers
- Enfileiramento por VC no adaptador de porta ATM IMA PA-A3-8T1/E1
- Configurando o enfileiramento ATM por VC no MC3810

Voz e Qualidade de Serviço (QoS)

P. Como funcionam a Fragmentação e Intercalação de Links (LFI)?

A. O tráfego interativo, como Telnet e Voice over IP, é suscetível para o aumento de latência quando a rede processa pacotes grandes, como transferências do Protocolo FTP (Protocolo de transferência de arquivo) em uma WAN. O retardo de pacote para o tráfego interativo é significativo quando os pacotes FTP são enfileirados em enlaces de WAN mais lentos. Foi criado um método para fragmentar pacotes maiores e enfileirar os pacotes (de voz) menores entre os fragmentos dos pacotes (de FTP) maiores. Os roteadores do Cisco IOS suportam vários mecanismos de fragmentação da camada 2. Para obter mais informações, consulte os seguintes documentos:

- REsumo dos mecanismos de eficiência de link
- VoIP sobre Frame Relay com Qualidade de Serviço (fragmentação, formatação de tráfego, IP RTP Priority)
- Links de VoIP por PPP com qualidade de serviço (LLQ / prioridade IP RTP, LFI, cRTP)

P. Quais ferramentas posso usar para monitorar o desempenho do Voice over IP?

A. Atualmente, a Cisco oferece várias opções para monitoramento da Qualidade de serviço (QoS) nas redes que usam as soluções de Voice over IP da Cisco. Essas soluções não avaliam a qualidade de voz por meio da PSQM (Medida perceptual de qualidade de voz), nem de alguns outros novos algoritmos propostos para esse tipo de avaliação. As ferramentas de Agilent (HP) e NetIQ estão disponíveis para essa finalidade. No entanto, a Cisco oferece ferramentas que fornecem uma ideia da qualidade de voz que você recebe, medindo atraso, instabilidade e perda de pacotes. Para obter mais informações, consulte Como usar o Agente de garantia de serviços e o Monitor de desempenho de redes interconectadas da Cisco para gerenciar a Qualidade de serviço nas redes de Voice over IP.

P. %SW_MGR-3-CM_ERROR_FEATURE_CLASS: Erro do recurso de gerenciador de conexão: Classe SSS: (QoS) - erro de instalação, ignorar.

A. O erro de instalação do recurso observado é um comportamento esperado quando uma configuração inválida é aplicada a um modelo. Ele indica que a política de serviço não foi aplicada devido a um conflito. Em geral, você não deve configurar a modelagem em class-default da política filial nos mapas de política hierárquicos. Em vez disso, configure-a na política matriz da interface. Essa mensagem, juntamente com o rastreamento, é impressa como consequência.

Com as políticas baseadas na sessão, a modelagem em class-default deve ser executada somente na subinterface ou no PVC. A modelagem na interface física não é compatível. Se a configuração for realizada na interface física, a ocorrência dessa mensagem de erro será um comportamento esperado.

No caso do LNS, outro motivo pode ser que a política de serviços possa ser provisionada por meio do servidor radius quando as sessões forem ativadas. Execute o comando **show tech para exibir a configuração do servidor radius e as políticas de serviço ilegais que estão instaladas por meio do servidor radius quando a sessão for ativado ou oscilar.**

Informações Relacionadas

- Conceitos básicos de ajuste de desempenho
- Suporte da Qualidade de serviço (QoS)
- Suporte Técnico e Documentação - Cisco Systems