UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

Qualidade de Serviço em redes IP
Serviços Diferenciados

por

FÁBIO RIBAS ARDEOLA

Dissertação submetida à avaliação como requisito parcial
para a obtenção do grau de Mestre em Ciência da Computação

Prof[a]. Liane Margarida Rockenbach Tarouco
Orientadora

Porto Alegre, outubro de 2001.

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Ardeola, Fabio Ribas

Qualidade de serviço em redes IP – Serviços Diferenciados / por Fabio Ribas Ardeola. – Porto Alegre: PPGC da UFRGS, 2001.

128 f.:il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2001. Orientadora: Tarouco, Liane Margarida Rockenbach.

1. DiffServ. 2. QoS. 3. Qbone. 4. Bandwidth Broker. 5. SLS, TCS e SLA. 6. PHB.

I. Tarouco, Liane Margarida Rockenbach. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitoria: Prof.ª Wrana Panizzi
Pró-Reitor de Ensino: Prof. José Carlos Ferraz Hennemann
Pró-Reitor Adjunto de Pós-Graduação: Prof. Philippe Olivier Alexandre Navaux
Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux
Coordenadora do PPGC: Prof. Carlos Alberto Heuser
Bibliotecária – Chafe do Instituto de Informática: Beatriz Regina Bastos Haro

Sumário

Lista de Figuras

Lista de Tabelas

Resumo

Este trabalho tem como finalidade apresentar uma visão geral dos assuntos relacionados à definição, configuração e gerenciamento dos serviços habilitados pela arquitetura de Serviços Diferenciados em relação à priorização de tráfego.

Nele, descrevemos as necessidades por QoS nas redes atuais e futuras, bem como o direcionamento da Internet 2 em relação ao QoS, mais especificamente voltada ao mecanismo DiffServ. Também são destacados os mecanismos de QoS para redes IP do fornecedor Cisco Systems, o qual possui a maior gama de mecanismos já implantados pelos fornecedores. Por fim, identificamos as características necessárias para participar do QBone, um testbed para Serviços Diferenciados.

Palavra-Chave: DiffServ, QoS, PHB, Bandwidth Broker, SLS, TCS, SLA e Qbone.

TITLE: “QUALITY OF SERVICE IN IP NETWORKS – DIFFERENTIATED SERVICES”.

Abstract

The present work provides a description of issues related to the definition, configuration and management of services enabled by the differentiated services architecture.

We will describe in this document informations like: the requirements for QoS in the Internet 2, the mecanisms implemtented by Cisco Systems for QoS and the architectural required for the QBone – an interdomain testbed for Differentiated Services.

Keywords: DiffServ, QoS, PHB, Bandwidth Broker, SLS, TCS, SLA and Qbone.

1 Introdução

A necessidade de QoS na Internet é um fato. Até hoje, a Internet tem oferecido apenas serviço do tipo "***best-effort***", no qual todo o tráfego é tratado da mesma forma e o melhor esforço é empregado para entregá-lo. Enquanto este procedimento se mantiver, qualquer um conectado à Internet receberá o mesmo atendimento, o que nem sempre é o suficiente. Por outro lado, milhares de empresas gostariam de ter a opção de pagar por serviços do tipo premium-level, e Internet Service Provider (ISPs); de ter a habilidade de prover classes de serviços diferentes baseadas em aplicações e requisitos de usuários.

A questão é: qual é o melhor mecanismo para implementar QoS? A resposta tem sido elaborada por técnicos, políticos e ideologistas por quase vinte anos. Esforços como o ST-2, realizados pela BBN Corp. (uma divisão da GTE Internetworking), foram considerados tecnicamente visíveis. Entretanto, devido ao fato do ST-2 ser um método de QoS orientado à conexão, o qual requer que cada roteador ao longo da rota reserve os recursos requisitados – o que é quase impossível na Internet –, acabou sendo abandonado.

Na década de 1990, foram propostos outros modelos para Internet como: o Integrated Services e o Differentiated Service (DiffServ), os quais abordam vários tipos de serviços, incluindo serviços best-effort e real-time, além de permitir reserva de banda.

No caso do mecanismo DiffServ, ele é baseado no campo ***type-of-service*** (ToS) dentro do cabeçalho IPv4. Este campo inclui poucos bits, conhecidos como IP Precedence, que são utilizados para priorizar tráfego através de enfileiramento diferenciado nos roteadores. Por exemplo, tráfego de alta prioridade, indicado pelo valor mais alto do campo IP Precedence, deve ser colocado na fila de alta prioridade dentro no roteador e repassado antes dos pacotes das filas de baixa prioridade.

O DiffServ, por ser um mecanismo não orientado à conexão (cada pacote carrega sua própria informação do nível de serviço), segue o espírito do IP e tem a vantagem de ter sido desenvolvido na Internet. Uma vez utilizado, o DiffServ permite que usuários e provedores de serviços possam priorizar o tráfego de rede, bem como prover tráfego garantido.

O DiffServ é uma arquitetura que está sendo amplamente estudada e testada na Internet2, pelo Internet2 QoS Working Group, e tanto a comunidade Internet como os fabricantes de equipamentos estão apostando muito na sua padronização como solução para o tratamento de tráfego em redes IP.

Assim, uma vez que nosso objetivo é apresentar um estudo sobre Qualidade de Serviço em redes IP, e a arquitetura DiffServ tem muita chance de se tornar um padrão, direcionaremos o presente trabalho com vistas a apresentá-la como uma solução para a diferenciação de tráfego.

Aqui trataremos dos assuntos relacionados à definição, configuração e gerenciamento dos serviços habilitados pela arquitetura de Serviços Diferenciados. Em relação à

especificação da sua arquitetura, apresentaremos um interdomain testbed, conhecido como Qbone.

O capítulo 2 apresenta uma visão geral sobre Qos justificando sua necessidade e apresentando uma introdução sobre os modelos de serviços Qos fim a fim.

No capítulo 3 é discutida a adoção do framework DiffServ (Differentiated Services) na Intenet2. Os requisitos de Qos e a dificuldade em atendê-los no contexto onde diversificadas administrações cooperam é apontado como uma dificuldade para a escalabilidade da solução DiffServ.

A arquitetura dos serviços diferenciados é apresentada no capítulo 4.

O capítulo 5 aborda os diferentes tipos de serviços candidatos a usar Serviços Diferenciados, classificando-os em função de seus requisitos.

No capítulo 6 é apresentado o modo de provisionamento e configuração de serviços diferenciados incluindo a distribuição de informação e o gerenciamento de recursos dentro do domínio.

No capítulo 7 é feita uma avaliação do DiffServ e são apresentados os quatro serviços propostos para uso na Internet2.

No capítulo 8 são apresentados mecanismos para implementar QoS, desenvolvidos pela CISCO.

No capitulo 9 é apresentada a arquitetura QBone,uma arquitetura de teste entre domínios para Serviços Diferenciados.

No capítulo 10 é apresentada uma solução adotada para assegurar Qualidade de Serviço seguindo-se um capítulo com algumas conclusões.

2 Visão Geral Sobre QoS

Qualidade de Serviço refere-se à habilidade da rede em prover melhores serviços a um tráfego de rede selecionado, sobre vários tipos de tecnologias, incluindo: Frame Relay, Asynchronous Transfer Mode (ATM), Ethernet, redes 802.1, SONET e redes IP. Em particular, características de QoS provêem mais e melhores serviços de redes, uma vez que:

- suportam bandwidth dedicado;
- possuem melhorias em relação à perda;
- tem mecanismos para evitar e gerenciar congestionamento de rede;
- tem mecanismos para conformação do tráfego da rede;
- tem mecanismos para configuração de priorização de tráfego através da rede.

Dentro da arquitetura de QoS, os seguintes componentes são necessários para assegurar QoS através de redes heterogêneas:

- QoS dentro de um único elemento de rede, o qual inclui características de enfileiramento, programação e conformação de tráfego;
- técnicas de sinalização QoS fim a fim entre os elementos de rede;
- funcionalidades de policiamento e gerenciamento de QoS para controlar e administrar tráfego fim a fim através da rede.

Nem todas as técnicas são apropriadas a todos os roteadores da rede, porque roteadores de borda e roteadores de backbone necessariamente não realizam as mesmas operações, assim, as tarefas de qualidade de serviços podem ser diferentes [CIS99].

Em geral, roteadores de borda realizam as seguintes funções de QoS:

- Classificação de pacotes
- Controle de admissão
- Gerenciamento de configuração

Em geral, roteadores de backbone realizam as seguintes funções de QoS:

- Gerenciamento de congestionamento
- Função de evita congestionamento

2.1 Quem necessita de QoS?

Todas as redes podem se beneficiar dos aspectos de QoS de modo a melhorar a eficiência da rede, independente se esta for uma pequena corporação, uma grande empresa ou um provedor de serviço Internet (ISP). Diferentes categorias de usuários de rede têm seus próprios requisitos de QoS; em muitas áreas, entretanto, os requisitos extrapolam.

Grandes empresas, por exemplo, necessitam prover soluções de QoS através de várias plataformas de rede; fornecendo soluções para plataformas heterogêneas, nas quais, na maioria das vezes, ocorre uma configuração de QoS para cada tipo de tecnologia envolvida. Uma vez que grandes redes implementam aplicações mais complexas e críticas, e possuem tráfego de Web em crescimento, utilizar QoS serve para priorizar este tráfego de modo a assegurar que os usuários alcancem seus requisitos [CIS99].

No caso de ISPs, a necessidade real é a de escalabilidade e performance. Por exemplo, ISPs que têm oferecido conexões do tipo best-effort IP agora estão também transferindo voz, vídeo e outras aplicações de dados em tempo real. O QoS responde à escalabilidade e performance necessárias para que estes ISPs possam distinguir os diferentes tipos de tráfego, além de habilitá-los a oferecer serviços diferenciados a seus clientes.

Por fim, pequenos e médios segmentos de negócios podem se utilizar do QoS para manipular a difícil tarefa de utilizar a cara conexão WAN para a Internet de maneira mais eficiente.

2.2 Por que empregar QoS?

Empregar QoS na rede provê [CIS99]:

- Controle sobre os recursos – controlar quais são os recursos que estão sendo utilizados (bandwidth, equipamentos, facilidades de wide-area, etc). Por exemplo, limitar a banda consumida sobre um backbone por uma transferência de File Transfer Protocol (FTP) ou dar prioridade a uma base de dados importante;

- Serviços particulares – no caso de um ISP, o controle e a visibilidade providos pelo QoS habilita o oferecimento de serviços diferenciados a seus clientes;

- Coexistência de aplicações de missão crítica – o QoS garante que a rede será utilizada eficientemente por aplicações de missão crítica; que o bandwidth e atrasos requisitados por aplicações sensíveis ao tempo estarão disponíveis; e que outras aplicações, utilizando o link, não afetarão o tráfego de missão crítica.

2.3 O modelos de serviços QoS fim a fim

Um modelo de serviço, também chamado nível de serviço, descreve um conjunto de características QoS fim a fim.

O QoS fim a fim é a habilidade da rede em assegurar requisitos de serviço para um tráfego de rede específico, de um fim da rede a outro. Encontramos três tipos de modelos de serviços QoS : best-effort, integrated e differentiated services.

Observação: os modelos de serviços QoS diferem uns dos outros em função da maneira como eles habilitam a aplicação no envio de dados, e o modo pelo qual a rede tenta enviar estes dados.

Consideremos os seguintes fatores quando houver de se decidir sobre qual tipo de serviço implementar na rede [CIS99]:

- A aplicação ou problema a ser resolvido. Cada um dos três tipos é adequado para certas aplicações;

- O tipo de habilidade ao alocar os recursos de rede;

- Análise custo-benefício. Por exemplo, o custo de implementar e desenvolver serviços diferenciados é certamente mais caro que o empregado em serviço best-effort.

Esta seção descreve estes modelos de serviços:

- Best-Effort Service
- Serviços Integrados
- Serviços Diferenciados

2.3.1 Best-Effort Service

O Best-effort é um modelo de serviço único no qual uma aplicação envia dados quando desejar, em qualquer quantidade e sem requisitar permissão ou informar primeiro a rede. Para serviços best-effort, a rede entrega os dados se ela puder, sem qualquer tipo de segurança de entrega, atraso associado ou throughput [SHE99].

2.3.2 Serviços Integrados

Serviços Integrados é um modelo de serviço múltiplo que acomoda múltiplos requisitos de QoS. Neste modelo, a aplicação requisita um específico tipo de serviço da rede antes de enviar os dados. A requisição é realizada através de sinalização; a aplicação informa a rede do seu perfil de tráfego e requisita um tipo particular de serviço [INS99].

A aplicação envia dados apenas depois que ela recebe a confirmação da rede. Envia dados de acordo com as regras descritas no perfil de tráfego. A rede realiza controle de admissão, baseada na informação da aplicação e recursos de rede disponíveis. Ela realiza a manutenção por estado do fluxo e então realiza classificação de pacotes, policiamento e enfileiramento inteligente baseado neste estado.

Características para prover serviços de carga controlada – *integrated service*:

- O Resource Reservation Protocol (RSVP) pode ser utilizado por aplicações para sinalizar seus pedidos de QoS para o roteador;
- Mecanismos de enfileiramento inteligente podem ser utilizados com o RSVP para prover os seguintes tipos de serviços:

— Serviços com taxa garantida: permite que aplicações reservem banda de modo a encontrar seus requisitos. Por exemplo, uma aplicação de voz sobre IP pode reservar 32 Mbps fim a fim usando RSVP mais Weighted Fair Queueing (WFQ);

— Serviços de carga controlada: permite que aplicações tenham pouco atraso e alto throughput, até durante o tempo de congestionamento. Por exemplo, aplicações do tipo playback podem usar RSVP com Weighted Random Early Detection (WRED).

2.3.3 Serviços Diferenciados

Serviços Diferenciados é um modelo de serviços múltiplos que pode satisfazer diferentes tipos de requisitos de QoS. Entretanto, diferentemente do modelo de Serviços Integrados, uma aplicação usando Serviços Diferenciados não sinaliza explicitamente o roteador antes de enviar o dado [INS99].

Para os serviços diferenciados, a rede tenta entregar um tipo particular de serviço, baseado na QoS específica de cada pacote. Esta especificação pode ocorrer de diferentes maneiras. Por exemplo: usando o bit IP Precedence setado em pacotes IP ou endereços de fonte e destino. A rede usa a especificação QoS para classificar, conformar e policiar tráfego, e para realizar enfileiramento inteligente.

O modelo de serviços diferenciados é usado por inúmeras aplicações de missão crítica e para prover QoS fim a fim. Tipicamente, este modelo de serviço é apropriado para fluxos agregados porque ele realiza classificação de tráfego.

Características do modelo de Serviços Diferenciados:

- O Committed Access Rate (CAR), realiza classificação de pacotes através do IP Precedence e conjuntos de QoS. O CAR realiza medições e policiamento de tráfego, provendo gerenciamento de banda;

- Esquemas inteligentes de enfileiramento tal como WRED e WFQ podem ser utilizados com CAR para entregar serviços diferenciados.

TABELA 1.1 - Comparação entre os modelos de serviços

Comparação	**Best Effort**	**Serviços Integrados**	**Serviços Diferenciados**
QoS garantido	Não	Sim	Sim
Escalabilidade	Sim	Não	Sim
Configuração	Não	Dinâmico Por sessão Fim a Fim	Estático Fluxos agregados Dentro do domínio

3 Internet 2 – A necessidade por QoS

O objetivo principal da nova Internet é suportar o avanço das novas aplicações de rede. Ao contrário do que deveria ser, muitas dessas aplicações não são visíveis na Internet atual, pelo fato do modelo presente de entrega “best-effort” não prover a mínima solicitação de performance fim a fim assegurada. Para habilitar estas aplicações, a nova Internet deve prover qualidade de serviço (QoS) funcional que permita às mesmas reservarem recursos de redes chaves sem causar impacto no tráfego best-effort [TEI99c].

Com o passar dos últimos anos, as aplicações Internet e a comunidade de engenheiros têm identificado um conjunto de requisitos para QoS na Internet, baseado em necessidades de aplicações e de engenheiros. Na verdade, o Internet2 QoS Working Group tem estudado estes requisitos e recomendado a adoção de Differentiated Services (DiffServ) para QoS.

O framework DiffServ tem emergido nos últimos anos como uma forma simples e escalável de QoS que provê serviços significativamente fim a fim através de múltiplas nuvens de rede administrativamente separadas, sem necessidade de complexidade. Estes objetivos são equivalentes aos da Internet2 QoS pois colocam a ênfase em simplicidade, escalabilidade, interoperabilidade e administrabilidade.

3.1 O problema na Internet

Na Internet atual, cada elemento ao longo do caminho do pacote IP não faz nada mais que o melhor esforço para entregar o pacote a seu destino. Se a fila do roteador é sobrecarregada, pacotes são descartados com pouca ou nenhuma distinção entre tráfego de baixa prioridade e tráfego urgente. Isto é conhecido como serviço best-effort [TEI99c].

Para funcionar corretamente, muitas aplicações avançadas requisitam o máximo de banda garantida e o mínimo de atraso do pacote (latência), os quais o best-effort não pode disponibilizar. Por exemplo, ferramentas remotas de colaboração geralmente têm requisitos baseados nos limiares de percepção humana que poderíamos traduzir como banda específica e latência necessária. A falha em atender a estas necessidades resulta em tornar a aplicação inútil.

Implementar QoS tem a finalidade de alcançar a coexistência do tráfego best-effort com o tráfego QoS, compartilhando recursos de rede, de modo que os benefícios de circuitos comutados de rede (com performance garantida) e rede best-effort (baixo custo) sejam simultaneamente alcançados.

3.2 Requisitos de QoS

De modo a alcançar QoS na Internet, o Internet2 QoS Working Group identificou alguns requisitos, são eles [TEI99c]:

- Habilitar aplicações avançadas;
- Escalabilidade;
- Administração;
- Medição;
- Admitir múltiplas e interoperáveis implementações de pedaços individuais de equipamentos e nuvens;
- Suporte de sistemas operacionais e middleware.

3.2.1 Aplicações avançadas

Quando perguntamos qual é a qualidade de serviço que necessitam da rede, desenvolvedores de aplicações tendem a sorrisos e dizem coisas como: “Eu necessito de toda a banda que puder ser dada com pouca latência, jitter e perda.” Esta resposta não é na verdade uma resposta séria. A realidade do desenvolvimento de aplicações de rede para best-effort tem levado muitos desenvolvedores de aplicações a se tornarem experts em escrever aplicações adaptadas, que funcionam corretamente sobre uma grande variedade de throuputs. Desenvolvedores estão conseqüentemente desacostumados em ponderar adequadamente os requisitos de suas aplicações [TEI99c].

Para suportar o desenvolvimento de aplicações ditas avançadas, o protocolo de transporte Internet – TCP – tem sido cuidadosamente projetado e aperfeiçoado nos últimos 25 anos em relação a congestionamento e largura de banda. Na Internet atual, novas conexões nunca são negadas pelos roteadores (a menos que não haja rota para o destino solicitado) e toda a performance da conexão cai se houver demasiado aumento na carga da rede. Não existem sinais de " sem linha" na rede com best-effort embora os servidores possam ser configurados para aceitar conexões até um limite estabelecido.

Em contraste, o usuário da rede com QoS habilitada perceberá um modelo de serviço muito parecido com o sistema de telefonia. Em primeiro lugar, ocorre o estabelecimento da conexão (call setup), na qual o usuário procura iniciar a conexão e reservar os recursos necessários. Então, assumindo que a chamada seja completada, é garantido ao usuário um canal de acordo com seus requisitos no qual ocorre a comunicação. Alternativamente, no momento de call setup, o usuário talvez receba um sinal ocupado e lhe seja negado o privilégio de conectar em um nível de QoS desejado.

Durante os últimos anos, investigações e discussões sobre as necessidades de algumas aplicações têm identificado requisitos específicos de rede para aplicações avançadas. Juntamente com os desenvolvedores, essa discussão revelou um consenso em relação aos tipos de parâmetros de transmissão que são relevantes.

Uma dimensão fundamental de qualquer requisito de QoS de aplicação é o conjunto de parâmetros de transmissão que são efetivamente necessários. Os parâmetros de transmissão mencionados como requisitos assegurados são bandwidth e latência. Para a nova geração de aplicações, estes requisitos de bandwidth vão de alguns megabits por segundo (< 10Mbps) a um máximo de latência de 39 a 500 mseg. Algumas aplicações requerem também jitter (variação no atraso dos pacotes), embora a maioria mascare o jitter através de buffers.

Uma importante classe de aplicações requer que uma QoS garantida seja aplicada, não apenas para fluxo unicast, mas para tráfego multicast. Pela complexidade da tecnologia multicast, o Working Group decidiu realizar uma extensão de QoS para fluxos multicast. Atuais interações de qualquer QoS focar-se-ão em fluxos unicast.

A maioria das aplicações tem requisitos específicos de transmissão, com thresholds difíceis para a percepção humana, e são altamente intolerantes a variações na performance da rede. Outras aplicações têm requisitos específicos, mas são tolerantes a descartes ocasionais na performance ou requer apenas que a média de performance num certo período de tempo se mantenha. Finalmente, existe o desejo por funcionalidade de rede que pode simplesmente tratar o tráfego de certas aplicações ou usuários “melhor” que outros tráfegos.

Por fim, existe um aspecto temporal para qualquer QoS garantida – quando a garantia começa e por quanto tempo ela vai permanecer? É o caso de aplicações como ensino a distância e controle de instrumentos remotamente. Nesses casos, a não reserva de conexão pode gerar problemas nas aplicações.

3.2.2 Escalabilidade

Certamente a grande descoberta da engenharia para prover QoS fim a fim por fluxo é a escalabilidade.

Existem três conseqüências na escalabilidade. Primeiro, o status apenas na borda sugere que a indicação de que se trata de um serviço especial /normal deve ser indicada no pacote. Segundo, a diversidade na administração e reenvio em altas velocidades devem concordar com uma semântica muito simples para a identificação. Terceiro, a abstenção do estado no centro significa que a rede vê apenas o agregado, o qual pode causar sérios problemas se o controle de compartilhamento não for finalizado com cuidado, embora todo o tráfego cruze por muitas associações administrativas. Para alcançar um serviço fim a fim, todos os domínios administrativos ao longo do caminho devem concordar com o tratamento do tráfego especial. Entretanto, acordos multilaterais dificilmente funcionam. Os ISPs são empresas competitivas que agem no seu próprio interesse e, quando necessário, contra os interesses de seus competidores. Para escalonarr a complexa administração de QoS entre domínios, serviços fim a fim devem ser construídos por concatenações de acordos bilaterais. Esses acordos não podem requerer realisticamente confiança estendida ou

controle através de associações administrativas e elas devem ser também isoladas de falha [TEI99c].

3.2.3 Administração

Como qualquer outro recurso, existe a necessidade de mecanismos para alocação e contabilização para QoS. Esses mecanismos devem operar eficientemente, provendo a usuários acesso rápido a características de QoS. Futuramente, tais mecanismos de administração deverão suportar um conjunto flexível de policiamento [TEI99c].

3.2.4 Medição

Desde que instituições e eventualmente usuários venham a pagar por serviços de QoS, deve haver maneiras com que o usuário possa medir e auditar a performance da rede. Os requisitos de medição implicam não apenas a necessidade de ferramentas de medição, mas também uma necessidade por métricas de performance de rede não entendidas. Provedores de rede podem necessitar de ferramentas de medição adicionais para apoiar o provisionamento e debug de serviços de QoS, e possivelmente suportar mecanismos de controle de admissão baseados em medidas automatizadas [TEI99c].

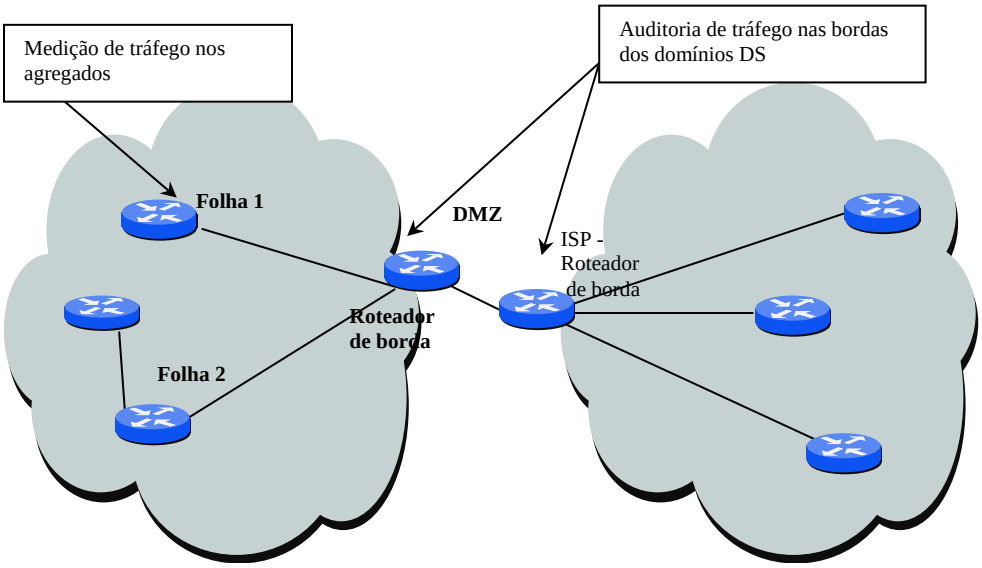

FIGURA 3.1 - Medição do tráfego

3.2.5 Admitindo múltiplas e interoperáveis implementações

Qualquer QoS deve admitir múltiplas e interoperáveis implementações de componentes chaves funcionais (por exemplo, elementos de packet forwarders, classifiers e admission control), bem como múltiplas e interoperáveis implementações de serviços de QoS através de nuvens individuais de rede.

3.2.5.1 Interoperabilidade de equipamentos

Qualquer tipo de QoS escolhido para ser implementado deve ser suportado por um ou por muitos fornecedores de equipamentos. Em redes heterogêneas grandes como a Internet, a interoperabilidade entre equipamentos de diferentes fabricantes é absolutamente essencial. Por sorte, a maioria dos fabricantes e provedores de serviços de rede entendem que a interoperabilidade baseada em padrões de tecnologias QoS é a chave para o sucesso de serviços fim a fim e para seu próprio sucesso financeiro. Para assegurar o sucesso do QoS e evitar isolamento técnico, a nova Internet deve seguir uma estratégia de QoS que siga os padrões de organizações de padrões para Internet tal como o Internet Engineering Task Force (IETF).

3.2.5.2 Interoperabilidade entre nuvens de rede

A habilitação QoS em fluxos e sinalização de call setup deveria ser tratada de maneira padronizada e bem entendida em ligações de nuvens com nuvens, mas nuvens devem permitir implementar QoS internamente de várias maneiras. Implementações internas de QoS podem variar dependendo das tecnologias empregadas na nuvem, policiamento interno e decisões de provisionamento [TEI99c].

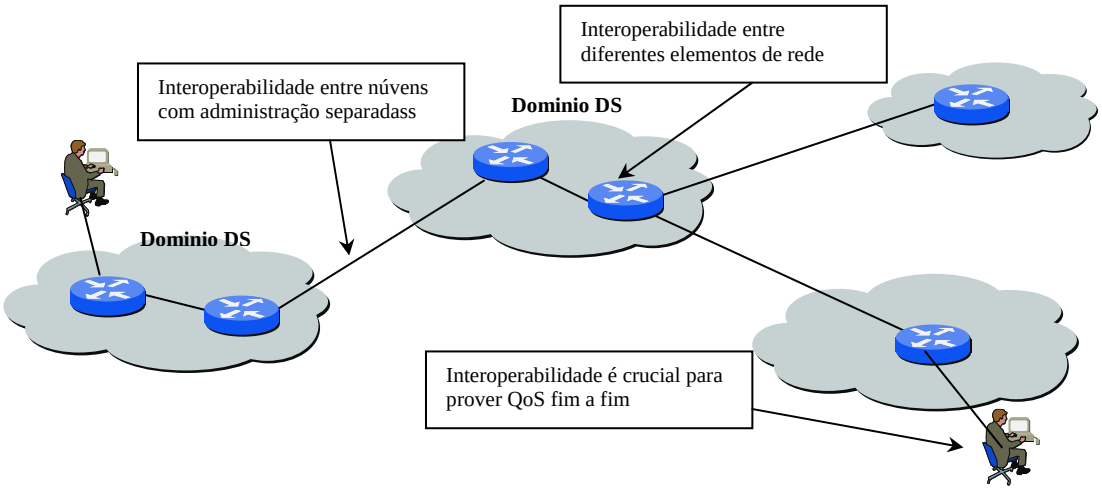

FIGURA 3.2.1 - Interoperabilidade entre nuvens

Esta é uma situação típica de ambiente na nova Internet. Como as interconexões são separadas por diferentes controles administrativos, existe a necessidade de padronizar a noção de QoS através das nuvens; só assim uma seqüência de nuvens poderá prover serviços fim a fim com QoS.

3.2.6 Suporte de Sistemas Operacionais

Eventualmente, os hosts devem estar aptos a inicializar requisições de QoS para seus fluxos. Em certos contextos, entretanto, serviços QoS podem necessitar de configurações estáticas. Os hosts também devem estar aptos a identificar apropriadamente os usuários para a rede no caso de autenticação, autorização e contabilização. Adicionalmente, para prover QoS fim a fim verdadeiro, sistemas operacionais de rede terão que suportar fluxos de QoS. Este tipo de funcionalidade em tempo real não está presente nos sistemas operacionais atuais, nos quais pacotes podem experimentar gargalos dentro da pilha da rede, o sistema de memória ou até schedule de processos [TEI99c].

4 Serviços Diferenciados

A arquitetura de serviços diferenciados procura prover um espectro de serviços na Internet sem ter que manter estados de fluxos para cada roteador. Isso ocorre através da união de fluxos dentro de um pequeno número de agregados nos quais são dados um pequeno número de tratamento diferenciado dentro da rede. O Diffserv elimina a necessidade de reconhecimento e armazenamento de informações sobre cada fluxo individual no roteador do core, alcançando assim a escalabilidade. O truque essencial é combinar um pequeno número de formas simples de tratamentos de pacotes agregados com um grande número de políticas de policiamento por fluxo para prover uma série de serviços [TEI99c].

Cada fluxo é policiado e marcado no primeiro roteador downstream confiável. Isso acontece de acordo com o perfil de serviço contratado, geralmente um filtro do tipo token bucket. Na visão do administrador de rede, o primeiro roteador downstream confiável é um roteador-folha na periferia da rede. O downstream do roteador folha mais próximo, em fluxo diffserv é unido com tráfego diffserv similar dentro de um agregado. Todo o reenvio e policiamento é realizado no agregado.

Outro benefício importante em relação a manipular tráfego agregado é a simplicidade necessária em relação aos negócios para construir serviços fim a fim. No modelo Diffserv, nuvens de rede individuais realizam contratos com nuvens vizinhas para prover contratos de serviços diferenciados para diferentes tipos de agregados. Como os contratos por fluxo, os contratos por agregado são caracterizados por perfis, de modo a forçar o contrato de tráfego agregado entre nuvens e garantir que novas chamadas que venham a exceder a capacidade do tráfego agregado não sejam admitidas. A arquitetura Diffserv provê administração separada para cada nuvem. Futuramente, uma vez que o contrato de agregados exista apenas na ligação em nuvens, o resultado é um conjunto de simples acordos de níveis de serviço bilaterais [TEI99c].

Em soma ao reenvio de pacotes no Diffserv, os componentes necessários incluem classificadores, policiadores, marcadores e um novo tipo de componente de rede conhecido como bandwidth broker.

O policiamento e marcação é realizado pelo primeiro roteador downstream do host que enviou. Quando uma decisão de controle de admissão local for feita pela nuvem emissora, o roteador-folha é configurado com o perfil de contrato do fluxo do serviço. No downstream do primeiro roteador-folha, todo o tráfego é tratado como agregado.

Na nuvem de entrada, o tráfego entrante é classificado pelos bits do per-hop behavior (PHB) dentro do agregado, os quais são policiados de acordo com o perfil do agregado. Dependendo do modelo de serviço Diffserv em questão, pacotes fora do perfil são descartados na borda ou marcados com um diferente PHB. Como o perfil de tráfego passa pela nuvem, ele pode passar por experiências de rajadas causadas por efeitos de enfileiramento ou aumento do agregado. Conseqüentemente, a nuvem pode necessitar

conformar a saída para prevenir que o tráfego seja policiado erroneamente na próxima nuvem downstream.

Finalmente, para realizar decisões apropriadas de controle de admissão interna e externa e para configurar folhas e dispositivos policiais de borda corretamente, cada nuvem é ajeitada por um ***bandwidth broker*** (BB). Quando um enviador sinaliza ao seu bandwidth broker local a inicialização de uma conexão, o usuário é autenticado e sua requisição é submetida a decisões locais de controle de admissão baseada em policiamento. Na representação do enviador, o BB então inicia um call-setup fim a fim ao longo da conexão da qual o bandwidth broker representa a nuvem. A abstração do bandwidth broker é importante porque ele permite a administração da nuvem separadamente.

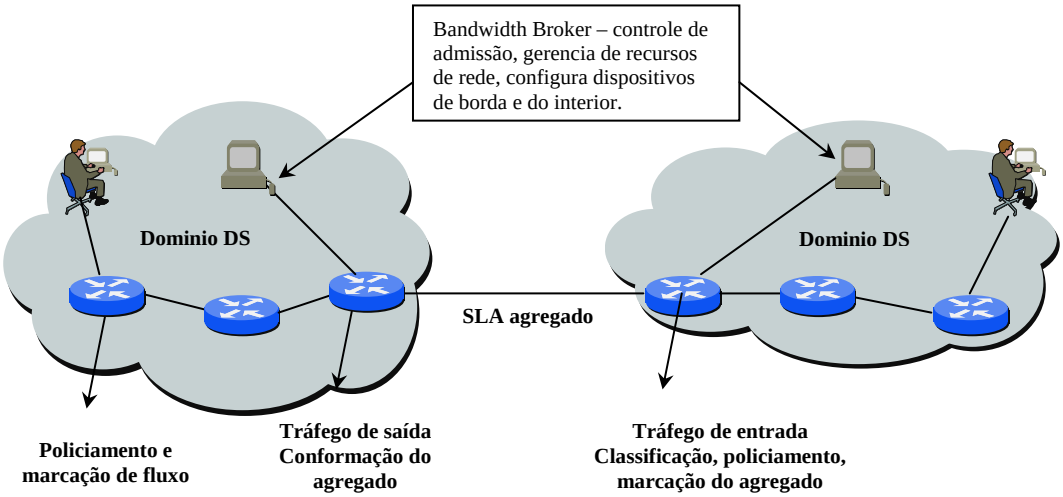

FIGURA 4.1 - Visão abstrata do DiffServ

4.1 A arquitetura do modelo de Serviços Diferenciados

Um serviço é definido como sendo “o tratamento global de um subconjunto definido de tráfego de um cliente, dentro de um domínio DS, ou fim a fim” [NIC99]. Ele define características significativas na transmissão de um pacote através de um ou mais caminhos dentro da rede. Essas características podem ser especificadas em termos quantitativos ou estáticos de throughput, atraso, jitter e/ou perda, ou pode ser especificado em termos de alguma prioridade de acesso a recursos de rede [BLA99].

A arquitetura de serviços diferenciados é baseada em um modelo simples no qual o tráfego que entra na rede é classificado, possivelmente condicionado na borda da rede e atribuído a diferentes agregações de comportamento. Cada agregação de comportamento é definida por um único DS codepoint. Dentro do core da rede, os pacotes são encaminhados de acordo com o *per-hop behavior* associado com o DS codepoint .

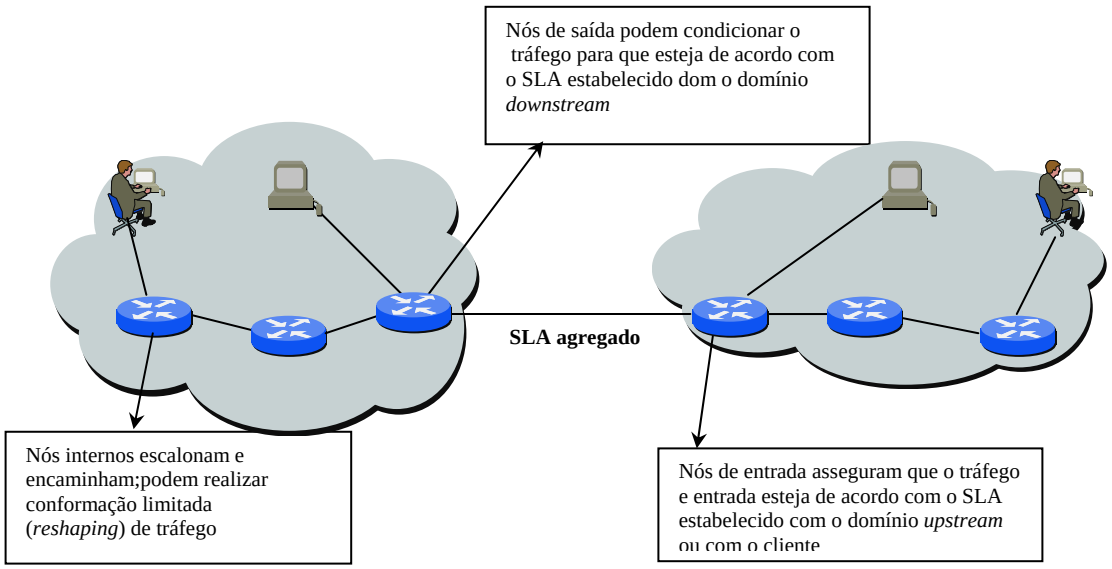

FIGURA 4.1.1 - Arquitetura física de Serviços Diferenciados

Essa arquitetura é composta de vários elementos funcionais implementados em nodos de rede, incluindo um pequeno conjunto de *per-hop behaviors*, funções de classificação de pacotes e funções de condicionamento de tráfego incluindo: medição, marcação, conformação e policiamento. Essa arquitetura alcança escalabilidade implementando classificação complexa e funções de condicionamento apenas nos nodos de borda de rede, e implementando *per-hop behaviors* para agregações de tráfego as quais são apropriadamente marcadas usando o campo DS. O *per-hop behaviors* (PHB) é definido para permitir uma média granular razoável de alocação de buffer e recursos de largura de banda em cada nodo [BLA99]. Assim, o PHB é responsável pelo tratamento de encaminhamento que os pacotes recebem nos roteadores. Tipos de PHBs [NIC99]:

- PHB EF (expedited Forwarding)
  - encaminhamento expresso (acelerado)
  - pouca perda, retardo e variação de retardo (jitter)
  - preferência total de encaminhamento

- PHB AF (Assured Forwarding)
  - grupo de PHBs de encaminhamento assegurado
  - 4 classes de serviços com três níveis de descarte
  - define tratamento diferenciado aos pacotes, do tipo “melhor que o melhor esforço - BBE”

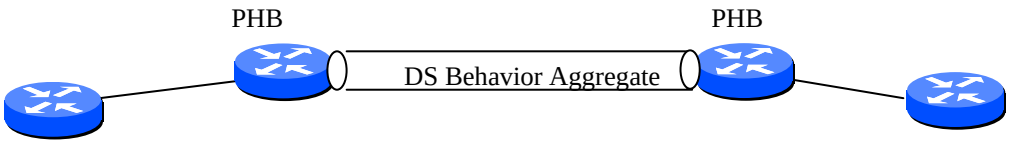

FIGURA 4.1.2 - Comportamento de encaminhamento

Mesmo sendo os PHBs que estejam no centro da arquitetura DS, é o serviço obtido pela marcação do tráfego que tem importância para o usuário final. Assim, poderíamos comparar os PHBs como sendo meros blocos de construção para os serviços. Os provedores de serviços combinam implementação de PHB com condicionadores de tráfego, estratégias de provisionamento e modelos de billing, como os responsáveis em habilitar o oferecimento de serviços para o usuário. Esses serviços são entregues após uma negociação de contrato entre o provedor e o cliente, respeitando os serviços a serem providos. Este contrato é conhecido como *Service Level Agreements* (SLAs). Muitos dos aspectos do SLAs (como termos de pagamento) estão além do escopo deste trabalho; como escopo, teremos um subset do SLA o qual provê as especificações técnicas do serviço. Ele é identificado como *Service Level Specification* (SLS) [BER99].

Ao oferecer um serviço é importante ter em mente que em um domínio DS devem ser consideradas as seguintes regras:

- os serviços do tipo DS são todos para tráfego unidirecional
- os serviços do tipo DS são tráfegos agregados

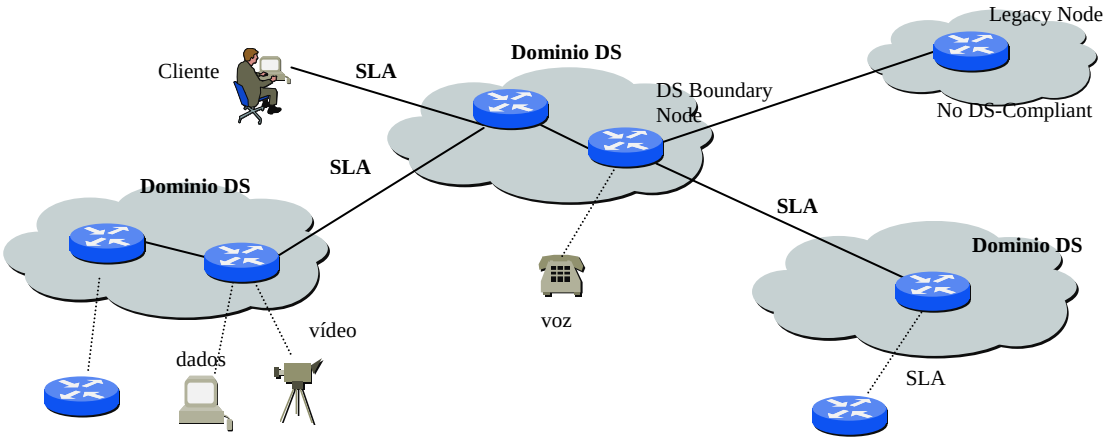

FIGURA 4.1.3 - Arquitetura lógica de Serviços Diferenciados

Um domínio DS é um conjunto contíguo de nodos DS os quais operam com um serviço de provisionamento de policiamento comum e um conjunto de grupos de PHB implementados em cada nodo. Um domínio DS normalmente consiste de uma ou mais redes dentro de uma mesma administração. A administração de um domínio tem a responsabilidade de assegurar que recursos adequados sejam provisionados e/ou reservados para suportar o SLA oferecido pelo domínio [BLA99].

4.2 SLSs e TCSs

Para cada serviço, diferentes aspectos técnicos do serviço a ser provido são definidos em forma de um *Service Level Specification* (SLS) que especifica todas as características e a performance esperadas pelo cliente. Devido ao fato dos serviços DS serem unidirecionais, as duas direções de fluxo devem ser consideradas separadamente. Um subset importante do SLS é o "*Traffic Conditioning Specification*", ou TCS.

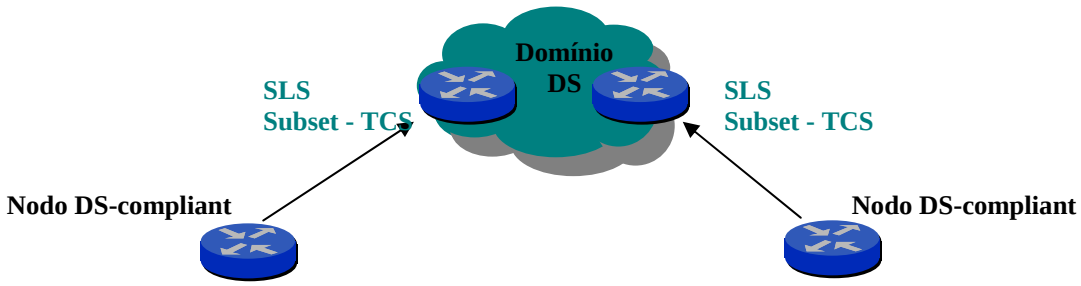

FIGURA 4.2.1 - SLS e TCS

O TCS especifica perfis de tráfego e ações para pacotes dentro do perfil (in-profile) e fora do perfil (out-of-profile). Perfis de tráfego são responsáveis em especificar regras para classificar e medir um fluxo, identificar quais são elegíveis e definir regras para determinar se um pacote está dentro ou fora do perfil. Um pacote dentro do perfil pode ser adicionado a uma agregação de comportamento diretamente, enquanto um pacote fora do perfil pode ser conformado antes da entrega da seguinte forma [NIC99]:

- pode ser atrasado até que esteja dentro do perfil
- pode ser descartado

Entre os parâmetros de serviço para cada nível de serviço que o TCS especifica temos [BER99]:

- parâmetros de performance, tais como: throughput, probabilidade de descarte e latência;
- indicação do escopo de cada serviço nos pontos de ingresso e saída;
- perfis de tráfego;

- disposição do tráfego submetido em excesso ao perfil especificado;
- marcação do serviço proporcionado;
- conformação do serviço proporcionado.

Em soma aos detalhes do TCS, o SLS pode especificar mais características de serviços, tais como:

- availability/reliability, as quais podem incluir comportamento no caso de eventos de falhas, resultando em re-roteamento de tráfego;
- serviços encriptados;
- roteamento;
- mecanismos de autenticação;
- mecanismos de monitoração e auditoria de tráfego;
- responsabilidades tais como: localização e funcionalidades de equipamentos, ações no caso de quebra de contrato, support capabilities.
- mecanismos de pricing e billing.

4.3 Serviços quantitativos e qualitativos

A arquitetura de Serviços Diferenciados pode suportar uma grande variedade de diferentes tipos de serviço. Classificar esses serviços significa associar um SLS a um serviço respectivo.

Alguns serviços podem ser claramente classificados como qualitativos ou quantitativos, dependendo do tipo de parâmetros de performance oferecidos.

Serviços qualitativos são aqueles que oferecem garantias relativas que somente podem ser avaliadas por comparação [NIC99].

Como exemplo de serviços qualitativos temos:

- o tráfego oferecido no nível de serviço A será entregue com baixa latência;
- o tráfego oferecido no nível de serviço B será entregue com baixa perda.

Serviços quantitativos são aqueles que oferecem garantias concretas que podem ser avaliadas por medições convenientes, independentes de outros serviços [NIC99].

Como exemplos de serviços quantitativos temos:

- 90% do tráfego entregue dentro do perfil no nível de serviço C não terá mais do que 50ms de latência;
- 95% do tráfego entregue dentro do perfil no nível D será efetivamente entregue.

Como serviços que possuem quantificação relativa temos:

- o tráfego oferecido no nível de serviço E terá duas vezes mais banda do que o nível F;
- o tráfego com drop precedence AF12 tem uma prioridade de entrega maior de que o tráfego com drop precedence AF13.

De uma forma geral, quando um provedor oferece um serviço quantitativo, será necessário especificar perfis de policiamento quantitativo. Em muito casos, perfis de policiamento quantitativo serão especificados até para serviços que não oferecem performance quantitativa [BER99].

4.4 O escopo do serviço

O escopo do serviço refere-se à topologia estendida sobre a qual o serviço é oferecido. Por exemplo, assumindo que um provedor oferece um serviço ao cliente que conecta a sua rede a um ponto de ingresso A. O serviço pode ser aplicado para:

- todo o tráfego do ponto de ingresso A para qualquer ponto de saída;
- todo o tráfego entre o ponto de ingresso A e o ponto de saída B;
- todo o tráfego do ponto de ingresso A para o conjunto de pontos de saída.

Os pontos de saída podem estar no mesmo domínio DS que os pontos de ingresso, ou podem estar em outros domínios que estão direta ou indiretamente conectados ao domínio de ingresso. Se o ponto de saída está em outro domínio, será necessário que o provedor de ingresso negocie o SLA com o próximo domínio peer, o qual recursivamente negociará com seus peers.

Em geral, os provedores estão aptos a oferecer serviços quantitativos mais eficientemente quando um conjunto específico de pontos de saída é especificado.

4.4.1 Serviços onde o escopo está amarrado ao receptor

Um caso especial de escopo de serviço é um serviço que controla todo o tráfego entre um ponto de ingresso e vários pontos de saída B. O SLS que define este serviço estará no ponto de saída B e poderá efetivamente permitir que o cliente controle a combinação de tráfego recebido do provedor. Mesmo sendo tal serviço teoricamente possível, ele compromete a especificação de Serviços Diferenciados a qual tem o objetivo de controlar qualidade no tráfego enviado, ao invés do recebido [BER99].

Um número de referências devem ser endereçadas por tal serviço, incluindo:

- tráfego direcionado ao ponto B proveniente do ponto de ingresso A, de acordo com os termos de SLS desse serviço; podem também ser controlados por um SLS para o

tráfego submetido no ponto A. Porém, os SLSs podem conflitar não sendo possível resolver os conflitos;

- estabelecer um perfil de tráfego para serviços onde o ingresso previna sobrecarga do receptor. Pode ser mais complexo do que outros escopos de serviço: perfis estáticos operam como se fossem ineficientes (dividindo o perfil de saída dentro de porções fixas) ou arriscados (permitindo ao ingresso enviar o perfil inteiro), já que perfis dinâmicos requerem processos e mecanismos de comunicação para coordenar as configurações. Por exemplo, as fontes talvez forneçam 1Mb/s enquanto o receptor pode receber apenas 9.6kb/s;

- sem perfis de ingresso efetivos para o serviço, negar o serviço será um problema sério.

Algumas das características de serviços orientados ao receptor podem ser providas por policiamento local e através de SLSs para o domínio onde o tráfego é enviado via ponto de saída, como descrito a posteriori.

4.5 SLS dinâmico vs. estático

Os SLSs podem ser estáticos ou dinâmicos. Os SLSs estáticos são a norma atualmente, e são um resultado da negociação entre o provedor e o cliente. Um SLS estático é definido por um acordo de data de início e pode ser periodicamente renegociado (em ordem de dias, semanas ou meses). Todavia, o SLS pode especificar que o nível de serviço mude em certas horas ao dia ou certos dias na semana, mas o contrato permanece estático.

O SLS dinâmico, ao contrário, pode mudar freqüentemente. Tais mudanças podem resultar, por exemplo, de variações na carga de tráfego oferecida, relativa a thresholds ou de mudanças no preço oferecida pelo provedor. Os SLSs dinâmicos mudam sem intervenção humana e requerem protocolos automatizados.

Os SLSs dinâmicos também apresentam problemas de mudanças para usuários e provedores de rede:

- provedores de rede têm que balancear freqüentemente mudanças de carga em diferentes rotas dentro da rede do provedor. Isso requer que o provedor adote mecanismos dinâmicos e automáticos;

- equipamentos de clientes terão que se adaptar a SLSs dinâmicos de modo a utilizar ao máximo as mudanças do SLS;

- aplicações de usuário final talvez tenham que adaptar seus comportamentos durante uma sessão para se utilizar do SLS.

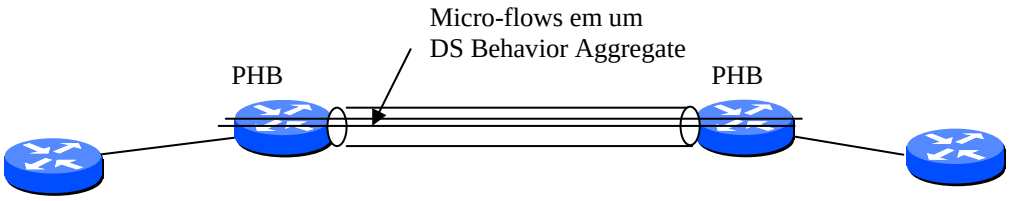

FIGURA 4.5.1 - Micro-flow de um agregado

É importante reiterar que os SLSs em Serviços Diferenciados se aplicam a tráfegos agregados e não a fluxos individuais. Para escalabilidade, não é necessário considerar modificações em um SLS toda vez que um micro-flow é somado ou removido do agregado.

4.6 Condições de provisionamento de tráfego em dispositivos de borda para provedores de serviços

Uma vez que um SLS tenha sido negociado, o provedor de serviço (e opcionalmente o cliente) irá configurar componentes de condicionamento de tráfego no limite das duas redes. Assim, o provedor de serviço tem como obrigação garantir os recursos ao cliente, porém sem que os recursos excedam os termos do TCS. E o cliente, ao contrário, tem como objetivo fazer o melhor uso do serviço adquirido do provedor [NIC99a].

Note que os próprios interesses do provedor requerem apenas que o provedor identifique:

- por qual nível de serviço o tráfego específico é submetido,
- por que cliente ele é submetido, e
- tráfego com SLSs de fins duplos (double-ended SLSs).

O tráfego do cliente talvez seja autenticado por conexões físicas através das quais ele chega ou por sofisticada criptografia. O provedor não necessita estar a par dos micro-flows do cliente para empregar Serviços Diferenciados.

Os quatro componentes de condicionamento de tráfego são identificados abaixo:

- Medidor
- Marcador
- Conformador
- Descartador

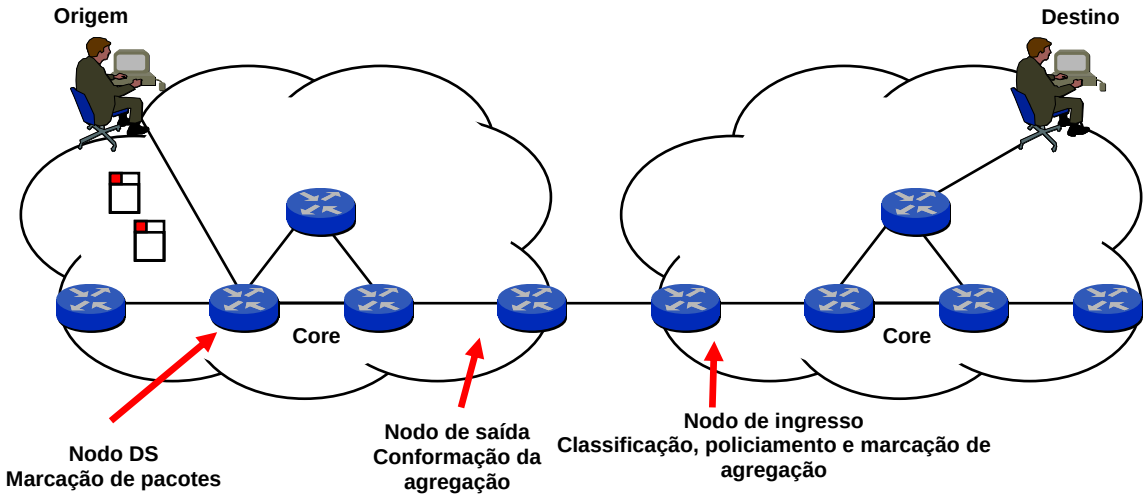

FIGURA 4.6.1 - Componentes do condicionamento

A combinação e interação de componentes de condicionamento de tráfego são selecionadas em bases de pacotes por pacotes pelo DS codepoint. Os parâmetros de configuração para componentes em cada codepoint são determinados por policiamento e perfis aplicados; dessa forma o condicionador policia o tráfego baseado no BA especificado pelo codepoint. Os medidores medem o tráfego submetido de acordo com o perfil de tráfego contratado (TCS), provendo controle de entrada para os outros componentes os quais implementam o policiamento:

- os conformadores policiam atrasando alguns ou todos os pacotes de uma seqüência de tráfego, de modo a levar o fluxo a tornar-se e acordo com o com o perfil de tráfego. Um conformador tem geralmente um tamanho de buffer finito e pacotes podem ser descartados se não houver espaço de buffer para assegurar o atraso dos pacotes [4];
- os descartadores policiam, descartando alguns ou todos os pacotes de uma seqüência de tráfego, de modo a levar o fluxo a ficar de acordo com com o perfil de tráfego. Este processo é conhecido como policiamento de fluxo. Note que o descartador pode ser implementado como um caso especial de um conformador, setando o tamanho do buffer para zero (ou quase) [BLA99];
- os marcadores policiam o tráfego remarcando o tráfego com um codepoint particular, somando o pacote a um comportamento DS particular. Isso ocorre:
  - mapeando codepoint/PHB específico do domínio;
  - rebaixando o fluxo fora do perfil de tráfego.

Em soma a estes quatro componentes, classificadores de tráfego são requisitados para separar o tráfego submetido dentro de diferentes classes. Os classificadores podem separar o tráfego baseado no campo DS (BA classifiers) ou podem fazê-lo baseado em múltiplos campos dentro do cabeçalho do pacote e até a partir do payload do pacote (MF classifier).

Os classificadores MF podem ser usados nos limites do provedor para certos serviços, para clientes do tipo per-micro-flow. Exemplos de tais serviços incluem: marcação ou conformação por fluxo. Na maioria da vezes, o tráfego irá chegar ao limite de um DS-domain pré-marcado e pré-modelado [BER99]. Entretanto, em interfaces com a rede do cliente sem DS, pode ou não ocorrer que o tráfego chegue marcado ou modelado.

Mesmo que o tráfego do cliente venha pré-marcado e pré-modelado, o provedor de serviço poderá realizar policiamento de tráfego no limite do ponto de ingresso, de modo a atender os próprios interesses do domínio. Isso pode resultar em que o tráfego seja remarcado ou descartado.

Condicionadores de tráfego podem ser encontrados dentro de um domínio DS, na borda de um domínio ou em um domínio não DS (non-DS domain) mas nem todos os quatro elementos precisam estar presentes em todos os nós de borda. Uma visão lógica de condicionamento de tráfego é mostrada na figura abaixo [NIC99a]:

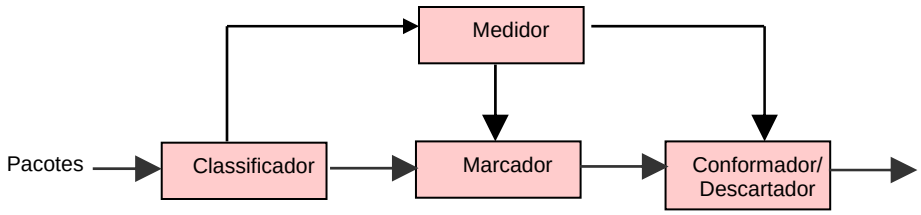

FIGURA 4.6.2 - Visão lógica de um classificador e condicionador de tráfego

Note que o condicionamento de tráfego pode não necessariamente conter todos os quatro elementos. Por exemplo, em casos onde não há um perfil de tráfego, pacotes podem apenas passar por um classificador e um marcador.

4.6.1 Funcionalidade mínima no ponto de ingresso do provedor

O provedor de serviços deve limitar o tráfego transportado em nome do cliente, de acordo com o TCS especificado. Um TCS simplificado pode ser representado em forma de tabela na qual cada linha tem o seguinte formato:

DS-Mark : Profile : Scope : Disposition of non-conforming traffic

A linha indica que o provedor compromete-se em transportar o tráfego marcado com *DS-Mark*, no nível de serviço correspondente, contanto que esteja conforme ao perfil. O tráfego que é submetido com *DS-Mark* o qual não está conforme ao perfil é submetido a *Disposition of non-conforming traffic*. Essa é geralmente uma ação de policiamento, tal como a remarcação para um nível mais baixo, atraso no conformador ou descarte. Alternativamente, ele pode ser transportado no nível de serviço requisitado, mas sujeito a mudanças [NIC99a].

Para prover essa funcionalidade mínima, os provedores devem configurar um classificador BA para separar o tráfego dentro de diferentes níveis de serviços requisitados, baseados no *DS-Mark*. Seguindo o classificador BA, cada classe deve ser medida de modo a estar de acordo com o perfil correspondente. Seguindo o perfil, tanto o descartador, como o conformador ou o re-marcador podem ser empregados.

4.6.2 Funcionalidade no ponto de ingresso do provedor para Double-ended SLSs

Para serviços quantitativos ou outros serviços que necessitem que double-ended SLSs (types 2 e 3) sejam implementados em domínios DS, eles devem especificar uma possível porta de saída para tráfego conforme o SLS. O condicionador de tráfego necessita considerar o endereço de destino dos pacotes como uma entrada para o processo de policiamento. Assim, esse tráfego não será aceito por portas de saída para as quais um SLS não existe. O serviço de linhas privadas virtuais é um exemplo de um serviço que utiliza esta funcionalidade.

Um QoS para VPN pode ser construído através de múltiplas instâncias do serviço tipo 2, uma malha de links QoS ponto a ponto. Já serviços do tipo 3 são mais usados para aplicações do tipo multicast.

4.6.3 Somando funcionalidades ao ponto de ingresso no provedor

As funcionalidades descritas nas duas seções anteriores servem apenas para proteger os recursos de rede do provedor nas linhas em termos de TCS. Não provêem assistência ao cliente. A marca dos pacotes e a modelagem de tráfego recaem inteiramente no cliente. Em alguns casos, o SLS pode ser utilizado pelo provedor, de modo a prover serviços adicionais para com o cliente. Tais serviços podem incluir:

- marcação de tráfego de específicos micro-flows para uma específica agregação de comportamento (marcação de DS-field);
- policiamento de tráfego de específicos micro-flows ou conjunto de micro-flows através de descarte ou conformação.

De modo a prover tal serviço, o provedor deve empregar um classificador MF em soma ao classificador BA. A necessidade de um classificador MF ocorre apenas quando o cliente requisita ao provedor uma forma de separação de tráfego ou autenticação em seu nome. O provedor pode implantar esse serviço dependendo da sua proporção de granulosidade e a quantidade de trabalho requisitado. Por exemplo, modelar milhares de clientes micro-flow talvez consuma recursos consideráveis em dispositivos de borda do provedor. Por outro lado, a marcação baseada em endereços de subnet origem irá consumir bem menos recursos.

4.6.4 Funcionalidade no ponto de saída do cliente

Na verdade, o cliente não necessita aplicar qualquer condicionamento de tráfego. Nesse caso, o cliente confia no provedor para a marcação definida na negociação com o classificador MF. Em muitos casos é preferível que o cliente marque. A marcação realizada pelo cliente será necessária em casos no qual o pacote do cliente está encriptado (como em casos de end-to-end IPSec). A marcação realizada pelo cliente habilita o direcionamento específico de tráfego de usuários ou aplicações para classes de serviço específicas. Isso seria difícil ou quase impossível de ser feito pelo provedor em nome do cliente, por exemplo, aplicações que usam portas voláteis e usuários que recebem endereçamento IP baseado em DHCP [BER99].

Em soma ao processo de marcação, existe o interesse pelo cliente na modelagem por nível de serviço na sua porta de saída. Isso ocorre por que o tráfego do cliente pode ser policiado por provedores com conseqüências indesejáveis (descarte de pacotes).

4.6.5 Funcionalidade no ponto de saída do provedor

No ponto de saída do domínio do provedor pode existir um SLA junto a um peer DS domain, o qual pode ser um provedor ou um domínio de usuário final. Nesse caso, o maior interesse do provedor é modelar o tráfego que deixa o domínio.

Dependendo do SLA, o ponto de saída pode requisitar a marcação e/ou policiamento ou modelagem do tráfego. Note que o tratamento de encaminhamento aplicado ao pacote no ponto de saída do domínio poderia ser selecionado pelo codepoint antes dele ser remarcado (caso contrário, o ponto de saída tem que suportar múltiplos codepoints para mapeamento de PHB).

Se o peer domain é um domínio non-DS o ponto de saída pode ser requisitado para remarcar todos os pacotes de modo a estarem conformes a um dos padrões de uso do byte TOS [NIC99a].

O provedor pode também desejar oferecer mais serviços ao cliente pelo policiamento de tráfego n a saída com granulosidade de micro-flow, se o cliente espera receber tráfego excessivo em um único BA e deseja aplicar um controle maior. Isso poderia ser especificado via uma SLS.

4.7 Provisionamento Interno

O provedor deve provisionar nodos internos na sua rede, de modo a encontrar as garantias oferecidas pelos SLSs negociados no limite da rede. Para fazer isso, o provedor pode usar mecanismos de condicionamento de tráfego similares aos usados no limite da rede. Entretanto, provedores são diferentes na aplicação de classificadores MF no interior

da rede. O provedor pode policiar periodicamente dentro da rede, por remodelagem, remarcação ou descarte de tráfego.

Os provedores de serviço são experientes em provisionamento de largas redes as quais oferecem serviços uniformes. Eles são equipados com ferramentas de prognósticos, ferramentas de modelagem de tráfego e medidores em tempo real. Em uma rede de serviços diferenciados, o provedor deve assegurar que os recursos garantidos ao tráfego de um nível de serviço não venha a comprometer a segurança independente do tráfego no outro nível de serviço. Como mencionado anteriormente, provisionamento interno em casos de SLSs dinâmicos terão a necessidade de protocolos de alocação de recursos dinamicamente.

4.8 Construção de serviços fim a fim

A arquitetura de serviços diferenciados propõe que um serviço fim a fim pode ser construído pela concatenação de serviços de domínios e SLAs associados ao cliente-provedor para cada um dos domínios onde o tráfego venha a passar.

Na verdade, como nem todos os PHBs e serviços podem ser concatenados; esta definição de serviços adequados e seus PHBs associados será o maior foco do futuro desenvolvimento de serviços diferenciados.

5 Tipos de Serviços - DS

Todo o tipo de transmissão pode ser tratada como dados. Uma vez que um sinal analógico é convertido para um sinal digital, ele pode ser tratado como se fosse um pedaço de dados. Entretanto, diferentes tipos de transmissão podem possuir diferentes tipos de requisitos.

Tanto voz como transmissões de vídeo de baixa qualidade apresentam alta tolerância a erros. Se um pacote ocasionalmente é descartado, a fidelidade na reprodução de voz e vídeo não será severamente afetada. Em contraste, pacotes de dados têm uma baixa tolerância a erros. Um bit errado pode mudar o significado dos dados.

Transmissão de voz, vídeo e dados também têm diferentes requisitos em relação a atrasos. Para que uma voz, que foi encapsulada em um pacote, possa ser traduzida com qualidade para um sinal analógico, o atraso de rede para estes pacotes deve ser constante e baixo. No caso de pacotes de dados, o atraso de rede pode variar consideravelmente. Pacotes de dados podem ser transmitidos de forma assíncrona através da rede, sem se importar com o tempo entre o emissor e o receptor. Em contraste, a transmissão de vídeo deve possuir uma relação de tempo entre o emissor e o receptor.

Pacotes de vídeo e voz, ocasionalmente, podem ser perdidos ou descartados. Em casos de eventos de excessivo atraso na rede, os pacotes podem ser descartados porque já não possuem utilidade. Essa perda não afeta severamente a fidelidade da voz, se a perda de pacotes for menor que 1% do total de pacotes transmitidos.

A transmissão de voz e vídeo também requer um tamanho de fila pequeno nos nodos da rede, de modo a reduzir o atraso e torná-lo previsível. Um tamanho de fila de pacotes de voz pequeno pode prevenir um overflow ocasional, o qual poderia resultar na perda de pacotes. Entretanto, pacotes de dados requerem uma fila de tamanho grande, de modo a prevenir que pacotes possam ser perdidos em condições de overflow.

Nessa seção, serão descritos exemplos de serviços e como eles podem ser suportados por específicos PHBs. Lembremos que tais exemplos têm caráter tão-somente ilustrativo, em se considerando a grande quantidade de serviços que podem ser empregados usando o modelo de serviços diferenciados.

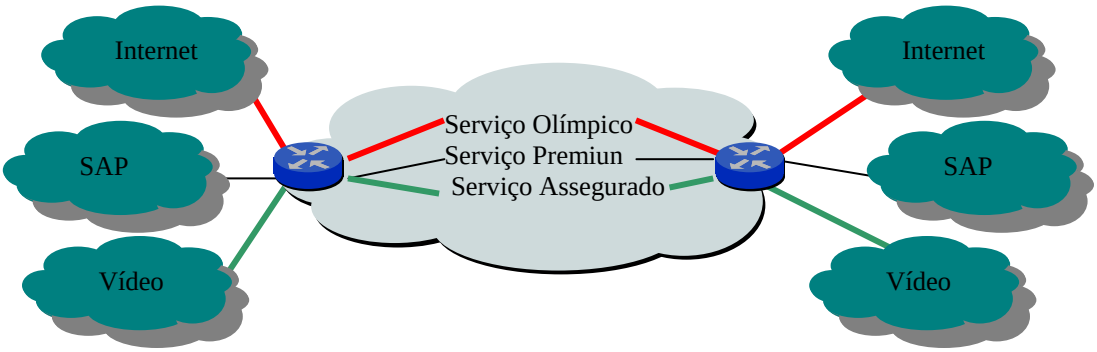

FIGURA 5.1 - Tipos de serviços DS

5.1 Melhor que Serviços Best-Effort (BBE)

Este é um serviço quantitativo que promete transportar tráfego de webservers em uma prioridade maior que a usada no método best-effort. Este tipo de serviço oferece perda de performance (não quantificada) relativa de um dado ponto de ingresso a qualquer outro ponto de saída. Também é conhecido como serviço do tipo Olímpico, no qual o contrato refere-se ao serviço "melhor", relativo a quem paga menos. Tem as seguintes características [NIC99a]:

- O PHB neste caso descarta as classes inferiores primeiro (AF);
- Tem como regra de policiamento, descartar ou remarcar pacotes fora de perfil;
- Também é chamado de Classe of Service;
- Tem como classes "olímpicas" de serviço de melhor esforço:
  1. Ouro
  2. Prata
  3. Bronze

5.1.1 Implementação do serviço

Neste exemplo, existe um SLS que define o serviço no ponto de ingresso do cliente. Este é o ponto no qual o cliente injeta respostas ao servidor web dentro da rede de serviços diferenciados. A informação no TCS pode ser representada da seguinte forma [BER99]:

AF11 Mark: 1 Mbps: Qualquer ponto de saída: Tráfego excessivo deve ser manipulado por marcação com marca AF13:.

Os pacotes submetidos para o serviço BBE devem ser marcados com o codepoint do campo DS correspondente a AF11 PHB. O provedor tem a promessa de transportar o tráfego de 1 Mbps do ponto de ingresso para qualquer ponto de saída a uma prioridade

maior que a do best-effort. Uma classe de serviço menor, correspondente a AF13 PHB, será aplicada ao tráfego submetido pela AF11 PHB, no excesso de 1 Mbps.

O provedor tem que prover um policiamento no ponto de ingresso. O tráfego submetido até o limite de 1 Mbps será direcionado ao AF11 PHB. O tráfego submetido em excesso ao 1 Mbps será remarcado para o AF13 PHB. Note-se que o esquema será preservado ordenando os pacotes desde que a AF11 e a AF13 utilizem uma única fila.

De modo a prover este serviço, o provedor terá que implementar a AF11 PHB e a AF13 PHB no equipamento do core da rede, que também deverá ser devidamente provisionado para recebê-las.

A AF11 PHB e a AF13 PHB podem ser implementadas, por exemplo, usando uma fila do tipo RIO. Provisionando parâmetros do tipo RED, por exemplo, o provedor está apto a controlar a prioridade do tráfego AF11 relativo ao tráfego AF13 em cada nodo da rede.

Uma vez que não existem garantias quantitativas, o provedor pode ser completamente liberal na sua estratégia de provisionamento e pode realizar ganhos de multiplexação estatística significativos. Também, a ausência de garantia quantitativa torna fácil prover tipos de serviço através de múltiplos domínios DS de provedores. Isso acontece porque não é necessário negociar.

5.2 Serviço de emulação de linhas privadas

Este é um serviço quantitativo que emula o serviço de linhas privadas tradicionais. Ele promete entregar o tráfego do cliente com pouca latência e pouquíssima probabilidade de descarte, até a taxa negociada. Acima desta taxa, o tráfego é descartado. Este tipo de serviço é encontrado tipicamente entre dois pontos específicos. Ele se encaixa para muitas aplicações de clientes. Entretanto, devido à alta garantia de qualidade ele acaba tendo um preço mais alto que serviços alternativos. Assim, ele acaba sendo utilizado apenas por aplicações que necessitam realmente deste tipo de serviço. Um exemplo de aplicação é a telefonia IP.

Este serviço também é conhecido como serviço do tipo Premium, no qual o contrato se refere à emulação de linha dedicada (VLL) a uma taxa de pico específica, e tem as seguintes características:

- O PHB neste caso tem como regra encaminhar o pacote primeiro (EF);
- Tem como regra de policiamento o descarte de pacotes fora de perfil;
- Na saída, os domínios devem conformar agregações EF para mascarar rajadas.

5.2.1 Implementação do serviço

Neste exemplo, consideramos um cliente com três redes geograficamente dispersas, interconectadas via um único provedor de rede. Os pontos de conexão do cliente serão

identificados como A, B e C. Em cada ponto conectado, um SLS descreve o serviço de linha privada a ser provido aos outros pontos. A tabela abaixo representa a informação requisitada no TCS da conexão do ponto A [BER99]:

EF-Mark: 100 Kbps: ponto de saída B: Descarte de tráfego não conforme
EF-Mark: 50 Kbps: ponto de saída C: Descarte de tráfego não conforme

Os pacotes submetidos pelo serviço de linha privada devem ser marcados com o codepoint no campo DS correspondente a EF PHB [EF]. Do ponto de ingresso A para o ponto de saída B, o provedor promete transportar até 100 kbps de tráfego. O tráfego excedente será descartado. Do ponto de ingresso A, para o ponto de saída C, o provedor promete transportar 50 Kbps de tráfego. É claro, existem algumas tolerâncias requisitadas em policiamento de tráfego como: jitter e tamanho de rajada. Entretanto, para serviços de linha privada o primeiro parâmetro de perfil de tráfego pode ser o sustained traffic rate.

O provedor provisionará policiamento no ponto de ingresso A para limitar o tráfego destinado ao ponto de saída B a 100 Kbps. Similarmente, um policiamento será configurado, de modo a limitar o tráfego destinado ao ponto de saída C a 50 Kbps. Esses policiamentos requerem classificação baseada no DS-mark e o endereço de destino em cada pacote.

A fim de prover este serviço, o provedor terá que implementar a EF PHB no equipamento do core da rede. A EF PHB pode ser implementada usando “strict priority queuing” ou alternativamente, aplicando pacotes marcados com EF no esquema WFQ (heavily weighted queue). O provedor terá que provisionar equipamentos no core da sua rede. Por exemplo, roteadores transportando tráfego entre o ponto A e ponto B e/ou C terão que ser provisionados considerando-se os recursos comprometidos pelo TCS no ponto A. Isso significa que um roteador o qual esta no caminho de A e B e de A e C, terá que ser considerado como tendo comprometido 150 Kbps de sua largura de banda como resultado do TCS colocado em A. Um roteador, apenas no caminho entre A e B, terá que ser considerado como tendo um comprometimento de 100 Kbps como resultado do TCS. É claro, o roteamento está sujeito a mudar, falhas nos caminhos podem também ser provisionadas. Para aumentar a segurança oferecida pelos serviços EF, os provedores podem empregar mecanismos de roteamento como: route pinning mechanisms ou QoS routing mechanisms.

5.3 Serviço quantitativo assegurado - Media Playback

Este serviço oferece menor garantia que o serviço de linha privada descrito acima, mas ele ainda é considerado um serviço quantitativo. Em particular, ele promete entrega de tráfego com um alto grau de confiança e com latência variável, porém limitada, até a taxa negociada. Acima desta taxa, o tráfego é sujeito a um atraso ou descarte significativo. Este tipo de serviço é tipicamente oferecido entre um conjunto específico de pontos e é empregado em muitas aplicações de clientes. Devido a sua variação de latência, ele acaba saindo mais em conta que o serviço de linha privada. Entretanto, devido ao seu limite de latência e alto grau de entrega, ele acaba tendo preço maior que outros serviços alternativos. Tal serviço é destinado particularmente a playback de áudio ou vídeo, no qual uma largura de banda considerável é necessária em bases contínuas, mas a natureza não interativa do tráfego torna-o um pouco tolerante a atrasos.

Também é conhecido como serviço do tipo Assegurado, no qual o contrato afirma que a rede parece estar "levemente carregada" para tráfego em perfil especificado (taxa e rajada), e tem as seguintes características [NIC99a]:

- O PHB neste caso descarta por último (AF);
- Tem como regra de policiamento remarcar pacotes fora do perfil para que tenham uma probabilidade de descarte maior;
- O tráfego em uma classe compartilha fila única;
- Na saída, os domínios podem também visualizar agregações AF

5.3.1 Implementação do serviço

Neste exemplo, consideramos novamente um cliente com três redes geograficamente dispersas interconectadas via um único provedor de rede. A tabela abaixo representa as informações requisitadas no TCS conectado ao ponto A [BER99]:

AF11-Mark: 100 Kbps sustained, 100 Kb bursts tolerated at up to 200 Kbps: Egress point B: Excess burst traffic over sustained rate marked with AF12-mark: Non-conforming traffic marked with AF13-mark: Max latency = 1 second

AF11-Mark: 50 Kbps sustained, 100 Kb bursts tolerated at up to 100 Kbps: Egress point C: Excess burst traffic over sustained rate marked with AF12-mark: Non-conforming traffic marked with AF13-mark: Max latency = 2 seconds

Os pacotes submetidos ao serviço de playback confiável devem ser marcados com o codepoint do campo DS correspondendo a AF11 PHB. Do ponto de ingresso A para o ponto de saída B, o provedor promete um transporte até 100 Kbps do tráfego tolerado (sustained traffic) com rajadas (burst) de 100 Kbps de tamanho e taxa de pico de 200 Kbps. Rajadas de tráfego excedentes serão marcadas com o codepoint AF12 e o tráfego fora do perfil será transportado com o codepoint AF13. Tão logo estas condições sejam

encontradas, a latência será limitada a um segundo. Note-se que para este serviço, o perfil de tráfego é descrito usando um conjunto completo de parâmetros de token bucket. Uma vez que o limite de latência para tal serviço é menos rigoroso que no serviço de linhas privadas, um certo grau de traffic burstiness pode ser tolerado.

O provedor deve suportar as AF11, AF12 e AF13 PHBs nos roteadores do core da rede. Estas PHBs podem ser providas, por exemplo, direcionando o tráfego marcado com AF11, AF12, AF13 para uma única fila RIO com alto limite de descarte. Os policiais na borda irão limitar a competição de tráfego na linha com o TCS, de modo a assegurar que a latência possa ser encontrada. Em soma, o provedor de serviço terá que provisionar dispositivos no core da rede. O provisionamento discutido em linhas privadas pode ser aplicado aqui, entretanto, em geral, o provedor de serviço tem a liberdade de ser menos conservativo no provisionamento e realizar melhores ganhos estatísticos.

5.4 Superposição de serviços quantitativos e qualitativos na mesma rede

O seguinte exemplo tem como objetivo apresentar uma rede com serviços quantitativos e qualitativos na mesma rede de serviços diferenciados. Um número de redes de *campus* corporativos seriam interconectadas por uma rede de serviços diferenciados, provendo serviços quantitativos entre os sites. Por exemplo, uma malha de serviços de linha privada poderia prover telefonia IP entre sites. Uma outra malha pode utilizar serviço de palyback de mídia usando o AF11 PHB, o qual habilitaria o playback de vídeo/áudio entre os sites. Em soma, cada site corporativo pode distribuir alguns níveis de serviço BBE para destinos arbitrários. Neste modelo, a rede de serviços diferenciados provê efetivamente uma malha de serviços quantitativos entre localizações fixas (similar a VPN). Esta malha é sobreposta em uma nuvem, suportando serviços BBE.

6 Provisionamento e configuração

O provisionamento de serviços diferenciados requer provisionamento e configuração cuidadosa.O provisionamento refere-se à determinação e alocação de recursos necessários em vários pontos na rede. O provisionamento pode: ditar a soma ou a remoção de recursos físicos em vários pontos (provisionamento físico); ditar a modificação de parâmetros operacionais dentro de equipamentos existentes na rede, de modo a alterar relativos compartilhamentos de recursos de rede os quais são alocados a uma ou outra classe de serviço (provisionamento lógico). A configuração refere-se à distribuição de parâmetros operacionais apropriados para equipamentos de rede de modo a alcançar objetivos de provisionamento.

Anteriormente, foi descrito brevemente requisitos de provisionamento e configuração, ambos na borda da rede e no interior da mesma. Nesta seção teremos como foco primeiro a coordenação de provisionamento e configuração através da rede, como estes serviços fim a fim podem ser providos com confiança. Discutiremos protocolos como: SNMP, CLI, RSVP, COPS e LDAP e no que condiz seu processo de provisionamento.

6.1 Provisionamento e configuração - Borda vs. Interior

De modo a ser breve, consideremos o termo ‘provisionamento’ como referência a provisionamento e configuração. O importante notar aqui é que provisionamento na borda da rede deve ser tratado separadamente de provisionamento no interior da rede. Desde que o provedor de serviços diferenciados vende um contrato (SLA) na borda da rede, podemos considerar o provisionamento de borda, o qual suporta SLSs, como sendo o responsável em determinar o provisionamento do interior do provedor. Por exemplo, um operador de rede não pode oferecer um SLS o qual não pode localizar recursos disponíveis no interior da rede. De uma forma geral, o processo geral de provisionamento interage entre bordas e interior. De agora em diante, referenciaremos o provisionamento em respeito a TCS em vez de SLS, já que o TCS é um componente do SLS que define detalhes de parâmetros de manipulação de tráfego.

6.1.1 Provisionamento de borda

Anteriormente, falamos do provisionamento mínimo que um provedor deve implementar para obter um TCS. Também foram discutidas as configurações adicionais que um provedor pode usar para prover serviços adicionais para o cliente. O último, na verdade, não está relacionado ao provisionamento de recursos dentro da rede de serviços diferenciados, mas assiste melhor o cliente determinando qual subconjunto de tráfego o qual o cliente faz uso neste contexto. Nesta seção, apenas consideramos o provisionamento mínimo requisitado na borda.

No mínimo, o provedor deve assegurar que recursos físicos suficientes estejam provisionados na borda de modo a poder atender os requisitos do TCS. Por exemplo, se a soma dos perfis suportados em um ponto de ingresso permitir 10 Mbps de tráfego, é inaceitável provisionar o link com um acesso T1. Um T3, entretanto, seria suficiente. Uma vez que o provisionamento físico é implementado, é necessário aplicar o provisionamento lógico apropriado. Isso é alcançado via configuração de policiamento que limita a quantidade de tráfego aceito pelo link T3, em cada nível de acesso e para duplos TCSs finais, para o ponto de saída apropriado. Também pode ser necessário configurar uma quantidade de buffer para as filas usadas para o serviço. O provisionamento similar é também apropriado em cada ponto de saída, se o agregado do perfil provisionado para a saída exceder a capacidade de saída do link.

6.1.2 Provisionamento Interno

Para o propósito de provisionamento do interior da rede é desejável entender ou controlar o volume de tráfego de cada classe a qual passa por cada nodo da rede. Pode-se entender o volume do tráfego passando por cada nodo se o tráfego for admitido de acordo com um TCS, o qual dita tanto o ponto de saída como o ponto de ingresso (este caso geralmente se aplica a serviços quantitativos e foram discutidos no contexto da EF PHB e serviços de linha privada). Enquanto volumes de tráfego não podem ser antecipados com 100% de exatidão a sua aproximação é possível especialmente com a ajuda do route pinning mechanisms.

Em outros casos, muitos (se não a maioria) dos serviços oferecidos por redes com serviços diferenciados não especificarão pontos de saída (como é o caso para serviços quantitativos) e não restringirão o tráfego submetido a pontos de saída específicos. Assim, nodos interiores terão que provisionar sem entender o volume de tráfego submetido por serviços quantitativos, os quais chegarão em cada nodo. É necessário, para ser capaz de provisionar redes de serviços diferenciados, suportar serviços quantitativos com pontos de saída específicos bem como serviços quantitativos, os quais não possuem pontos de saída específicos no mesmo recurso físico. Por fim, é necessário isolar o impacto do tráfego qualitativo em recursos reservados para tráfego quantitativo. Isto pode apenas ser encontrado se o primeiro é tratado com menor prioridade que o anterior. Assim, em geral, recursos terão que ser provisionados primeiro para tráfego quantitativo, usando mecanismos de provisionamento quantitativos. Então, o provisionamento quantitativo pode ser usado para alocar recursos restantes para tráfego qualitativo. O provisionamento qualitativo pode também ser aplicado para serviços os quais oferecem uma quantificação relativa de volumes de tráfego.

O impacto dos dois tipos de tráfego terá que ser isolado de modo a assegurar que eles não compartilhem codepoints PHB. Os PHBs usados para serviços quantitativos terão sempre alta prioridade de acesso, ao contrário dos usados para serviços qualitativos. Como resultado, é necessário policiar cuidadosamente o tráfego submetido a PHBs quantitativos. No caso de falhas pode ocorrer perda de tráfego de baixa prioridade. Em geral pode ser esperado que apenas uma pequena fração dos recursos em cada nodo seja provisionada para tráfego qualitativo.

Similarmente, uma fração significante de capacidade de tráfego deverá sobrar para serviços do tipo best-effort. Assim, se um congestionamento ocorrer ou se for necessário redirecionar o tráfego do tipo não best-effort em caso de um evento de falha, um destino ‘soft’ existirá para descarte de tráfego.

6.1.2.1 Provisionamento quantitativo do interior

Como discutido anteriormente, o provisionamento quantitativo é difícil. No entanto, os problemas daí decorrentes são passíveis de tratamento. Com conhecimento da topologia de roteamento da rede e os TCSs em cada borda, é possível calcular os recursos necessários em cada nodo interior para transportar o tráfego quantitativo oferecido nas bordas. Baseado em recursos de computação, os nodos interiores devem ser provisionados e configurados com capacidade suficiente para acomodar o tráfego quantitativo o qual chegará no nodo, enquanto deixa capacidade restante suficiente para acomodar certa quantidade de tráfego qualitativo.

O mecanismo de provisionamento descrito assume uma aproximação top-down, na qual o administrador de rede estuda a topologia da rede, o roteamento do tráfego, e calcula os requisitos de provisionamento. Uma aproximação alternativa usa sinalização para automatizar o processo. Por exemplo, mensagens do tipo RSVP poderiam ser lançadas ao longo do caminho que será seguido pelo tráfego submetido. Se um TCS deseja um serviço de linha privada a 100 Kbps, do ponto de ingresso A para o ponto de saída B, uma mensagem RSVP poderá ser transmitida do ponto A para o ponto B, com um flowspec especificando 100 Kbps. Esta mensagem passará por cada nodo nos quais recursos serão comprometidos. Em uma rede de serviços diferenciados, o RSVP poderia ser adaptado de modo a provisionar recursos requisitados pelo modelo de serviços diferenciados. Em uma rede que oferece um número estático de TCSs, tais mensagens de RSVP poderiam ser lançadas do TCS ponto de ingresso no momento que o TCS é inicializado, de modo a instantaneamente acumular provisionamentos em routers ao longo de vários routers. A vantagem deste tipo de aproximação é que ela não requer conhecimento explícito da topologia da rede.

Uma vez que os recursos requisitados para o tráfego quantitativo em cada nodo tenha sido determinado, o provisionamento do nodo consiste em instalar ou configurar interfaces com capacidade apropriada para facilmente acomodar o tráfego quantitativo que irá cruzar o nodo. Note-se que não há um significado preciso em ‘facilmente acomodar’. Um certo número de fatores devem ser considerados na determinação de capacidades apropriadas, de modo a dar um certo volume de tráfego quantitativo previsto. Isto inclui:

- Margem de erro
- Ganho estatístico desejado
- Capacidade restante para tráfego qualitativo

O primeiro, margem de erro, acomoda erros no seu cálculo, efeitos das mudanças passageiras de roteamento os quais na maioria das vezes são quantificados, efeitos de traffic clustering de como ele se move através da rede e outros. No caso do ganho estatístico desejado, ele refere-se ao grau o qual o provedor está disposto a apostar que nem todos os recursos do tráfego quantitativo estarão simultaneamente ativos nos limites ditados pelos TCSs no ponto de ingresso. Finalmente, o provedor deve determinar a capacidade que será reservada para tráfego qualitativo em cada nodo. Assim, se é determinado que 1 Mbps de tráfego quantitativo talvez passe por um nodo específico em uma direção específica, o provedor pode, então, instalar uma interface de 10 Mbps no nodo, de modo a servir ao tráfego correspondente. Isso deixaria 9 Mbps da capacidade completamente salva para tráfego qualitativo. Neste caso, o provedor assumiria que ganhos estáticos, que porventura venham a acontecer, seriam usados para compensar a margem de erro a qual poderia comprometer a disponibilidade do recurso.

Em soma, a instalação ou configuração de capacidade apropriada em cada interface, pode ser desejada uma configuração de policiamento, de modo a assegurar que o recurso atualmente consumido pelo tráfego quantitativo de mais alta prioridade não exceda as expectativas. Isso é especialmente importante se o provedor almeja alcançar um alto grau de ganho estatístico ou não pode ter uma margem de erro razoável. O policiamento não necessita ser configurado em cada nodo interior, mas deve provavelmente ser configurado em certos nodos chaves. Talvez, também seja necessário configurar recursos internos de roteadores (queues e buffers) para entregar o serviço requisitado.

6.1.2.2 Provisionamento qualitativo do Interior

Como explicado previamente, é necessário primeiro determinar os recursos que devem ser provisionados em cada nodo para tráfego quantitativo. Uma vez que estes tenham sido determinados, interfaces devem ser instaladas ou provisionadas para acomodar os recursos requisitados enquanto deixa capacidade suficiente para tráfego qualitativo. De modo a fazer isto, é necessário determinar os recursos requisitados no nodo para tráfego qualitativo. Desde que o tráfego qualitativo não pode ser assumido a um rota específica de fluxo, com o mesmo grau de predictabilidade – como o tráfego quantitativo –, esse problema de provisionamento é muito mais difícil e parâmetros de provisionamento devem ser estimados baseados em heurísticas, experiências e possivelmente em medidas em tempo real.

Uma vez que interfaces físicas tenha sido selecionadas, de modo a acomodar os recursos necessários pela carga do tráfego quantitativo computado e a carga de tráfego qualitativo estimado, uma configuração adicional é requisitada para suportar tráfego qualitativo. Tal configuração é importante devido à seleção de valores relativos (weights) para filas (queues) para diferentes níveis de serviços (em um esquema WFQ), ou à seleção de thresholds do tipo RIO ou RED, ou parâmetros de provisionamento lógico de recursos alternativos. É assumido que se o tráfego quantitativo é acomodado via mecanismos de filas similares (em oposição a strict priority queuing), os valores escolhidos para os parâmetros para o tráfego quantitativo efetivamente devem isolar o tráfego qualitativo efetivo.

Entretanto, os parâmetros de configuração que diferenciam os vários serviços qualitativos podem não prover um grau de isolamento entre os serviços qualitativos. Assim, pode ser necessário estimar o tráfego entrante para cada serviço qualitativo e antecipar a interação entre tráfego de diferentes serviços qualitativos. Pode ser impossível efetiva e conservativamente provisionar uma rede para certas combinações de serviços qualitativos. Para ajudar no provisionamento de um rede para serviços qualitativos, pode ser útil configurar policiamento para controlar o volume de tráfego entrante em um dado nodo.

Entretanto, tal policiamento talvez tenha que ser restrito a compartilhamento (no lugar do descarte) de modo a evitar violação de TCSs nas bordas.

6.2 Provisionamento estático vs. dinâmico

Por enquanto, nós consideramos apenas técnicas de provisionamento estático. Até o exemplo de RSVP usado para provisionamento assumia que as mensagens RSVP eram lançadas no momento que um TCS era iniciado em oposto ao método dinâmico. No caso de TCSs estáticos, o provisionamento é adequado para tráfego quantitativo. Entretanto, desde que o tráfego qualitativo oferece menos padrões prognosticáveis, ele age como se alterasse dinamicamente volumes de tráfego em diferentes nodos na rede, até quando o TCS é estático. Por esta razão, técnicas de provisionamento dinâmico são desejáveis e podem assistir os provedores de serviços de modo a fazer melhor uso dos recursos de rede. Em soma, o provisionamento dinâmico pode habilitar os provedores de serviços a um provisionamento mais liberal para serviços quantitativos, realizando ganhos estatísticos. Se considerarmos futuramente, em relação ao desejo de prover mudanças dinâmicas em TCSs, então, o apelo por técnicas de provisionamento dinâmico é muito forte.

O provisionamento dinâmico pode ser baseado em sinalização, baseado em medidas ou ambos. Por exemplo, o RSVP suporta sinalização baseada em provisionamento dinâmico. Os hosts sinalizam ao roteador a requisição de mais ou menos recursos e o roteador faz o ajuste de acordo. O hosts pode ou não submeter o tráfego na taxa a qual foi sinalizada, mas indiferentemente, os recursos são comprometidos. O provisionamento baseado em medidas pode ajustar os recursos comprometidos em resposta a carga de tráfego efetivamente medida no dispositivo. Enquanto serviços diferenciados não especifica qualquer forma de sinalização ou medição baseada em provisionamento, ambas podem ser utilizadas.

6.3 Distribuindo informações de configuração

O processo de provisionamento físico é, por necessidade, relativamente estático e não pode ser automatizado desde que requeira instalação de equipamentos físicos. Entretanto, o provisionamento lógico e configurações podem e devem ser automatizados. Nesta seção, abordaremos técnicas de distribuição de informações de configuração.

6.3.1 Distribuição top down de informações de configuração

No caso mais simples, os TCSs são estáticos e as bordas e o interior da rede são provisionados estaticamente através do processo de *pushing* da informação de configuração e *down* para o nodo de rede apropriado. A configuração dos nodos de borda requer primeiro o *pushing* da informação de policiamento de modo a forçar os TCSs. As informações de configuração para os nodos de borda são determinadas no momento que o TCS é negociado. Neste momento, os nodos são configurados pelo provedor. O administrador de rede pode usar um dos vários protocolos para fazer isto, incluindo SNMP ou CLI.

De modo a acomodar o tráfego submetido pelo provisionamento de um novo TCS, é necessário provisionar o interior da rede. Como discutido anteriormente, é possível computar o recurso requisitado para tráfego quantitativo. Assumindo que a capacidade física suficiente tenha sido provisionada, há configuração para logicamente provisionar a capacidade suficiente em cada interface do interior e para configurar policiamento para tráfego quantitativo em vários nodos interiores. Em soma, o provisionamento quantitativo requer a configuração de *queues*, WFQ *weights* e/ou parâmetros de RIO em vários nodos interiores, e pode também incluir a configuração de policiamento. Neste caso de configuração *top down* , as informações de configuração de interior é também *pushed down* via um protocolo de configuração tal como SMNP ou CLI.

A dificuldade de tal provisionamento *top down* é que ele requer um administrador de rede para coordenar o provisionamento de cada nodo da rede, nas bordas bem como no interior, de modo que a rede seja provisonisada fim a fim, em uma maneira consistente e habilitada a entregar eficientemente os serviços providos pelos TCSs. De modo a assistir o administrador de rede nessa tarefa, é útil considerar uma base de dados que mantenha informações da topologia corrente, bem como as TCSs na borda da rede. Essas informações são armazenadas em um formato ditado por um esquema padrão.

É claro, a base de dados é mantida de forma lógica centralizada – para fácil programação e modificação –, mas é fisicamente distribuída – para o caso de robustez e tolerância a falhas. Servidores de policiamento podem ser usados para extrair informações de base de dados e para convertê-las em informações de configuração, as quais são *pushed down* para nodos individuais. Neste cenário, os servidores de policiamento poderiam utilizar protocolos do tipo *diretory access protocol*, tal como LDAP para buscar informações do diretório e usar um protocolo de configuração como SMNP ou CLI para *push down* a informação de configuração para o nodo da rede. Note-se que é simples, o

servidor de policiamento e os esquemas de diretórios são uma comprovação do bandwidth broker. Em particular, o servidor de policiamento usa a qualidade de estar ciente da topologia da rede para provisionar nodos interiores, tal que certas rotas QoS fim a fim podem ser construídas e asseguradas pelos TCSs.

6.3.2 Distribuição de informações de configuração via sinalização

Um mecanismo alternativo de distribuição de informação de configuração é via transmissão de mensagens de sinalização entre nodos de borda do mesmo domínio de serviços diferenciados (*intra-domain signaling*). É interessante considerar a sinalização inter-domain, mas isso será tratado separadamente. Um exemplo de tal sinalização foi descrito previamente, no uso de um RSVP modificado. Tal sinalização é particularmente útil para o propósito de instalação de informações de configuração para serviços quantitativos os quais afetam especificamente caminhos e é um pouco mesmo útil (embora não inútil), para o propósito de configuração de serviços qualitativos. É como se tal sinalização fosse utilizada em conjunto, como provisionamento *top down*. Por exemplo, o esquema de diretório talvez indicasse a quantia de recursos para ser disponível para alta prioridade de serviços diferenciados em cada nodo. Estes limites são enviados para nodos individuais a priori. A sinalização proveniente da borda da rede, no momento da inicialização do TCS, poderia então ser usada para reivindicar recursos do pool de recursos quantitativos disponíveis em cada nodo. Alternativamente, nodos talvez consultem servidores de policiamento como os sinalizadores de requisições de recursos que chegam em cada nodo. O último modelo é similar ao uso da sinalização per-flow RSVP e PEP/PDP policy usada em rede RSVP tradicional. Informações de configuração quantitativa poderiam ser entregues no modo top down. A vantagem do último modelo é que o servidor de policiamento poderia ser dinamicamente atualizado com informações, com referência ao uso corrente dos recursos da rede. Neste modelo, é como se uma variedade de COPS fosse utilizada para comunicar-se entre nodos de rede e servidores de policiamento. Note-se que os COPS podem ser usados para distribuição de informações de configuração via *top down*, embora ele não tenha sido especificamente projetado para este propósito.

Uma das vantagens de configuração via sinalização é que ela facilita o suporte de TCSs dinâmicos. Os TCSs poderiam ser renegociados dinamicamente usando sinalização inter-domain. Tal renegociação poderia requisitar modificações dinâmicas a provisionamento dentro do domínio afetado, um processo que requer um protocolo de sinalização automatizado, tal como um agregado proveniente da sinalização RSVP entre nodos de borda de um provedor de domínio. Este protocolo poderia de fato representar um *bandwidth broker* distribuído para o domínio.

6.3.3 Modificações de base de informações de configuração em Real-Time Measurement

Um terceiro mecanismo para a configuração de nodos interiores poderia ser baseado em medidas da carga do tráfego corrente nos nodos chaves da rede. A configuração baseada em medidas é mesmo necessária para provisionamento quantitativo, desde que os padrões de tráfego quantitativo sejam relativamente previstos. Entretanto, ele pode aumentar significativamente a eficiência com a qual o provisionamento qualitativo pode ser alcançado. Por exemplo, nodos de rede podem suprir servidores de policiamento com medidas correntes de carga de tráfego qualitativos. Em resposta, *bandwidth broker* e servidores de policiamento talvez recalculem os relativos valores para filas de serviços diferenciados em um WFQ do nodo e entreguem a nova informação de configuração para o roteador. É como se a configuração, baseada em medidas para serviços qualitativos, pudesse ser usada em conjunto como se fossem configurações baseadas em sinalização para serviços quantitativos.

6.4 Implementando Bandwidth Broker

O objetivo final em relação a QoS é prover a usuários e aplicações uma alta qualidade na entrega de serviços de dados. No ponto de vista do roteador, o suporte à qualidade de serviço é dividida em três partes: definição de classes de tratamento de pacotes, especificação da quantidade de recursos para cada classe e classificação de todos os pacotes de entrada da rede dentro de suas classes correspondentes. O modelo DiffServ especifica a primeira e a terceira parte: ele especifica classes de tráfego, bem como provê um mecanismo simples de classificação de pacotes. Já o modelo Bandwidth Broker especifica a segunda parte, mantendo a informação de alocação atual do tráfego marcado e interpretando novas requisições.

O BB tem responsabilidades internas e externas referentes a gerenciamento de recursos e controle de tráfego. Internamente, um BB pode manter informações de requisições de QoS de usuários individuais e aplicações, e alocar recursos internos de acordo com regras de policiamento usadas para recursos específicos dentro do domínio. Externamente, o BB tem responsabilidades de configurar e manter acordos bilaterais de serviços com os BBs de domínios vizinhos de modo a assegurar a manipulação de QoS do tráfego de dados entre as bordas.

6.4.1 Bandwidth Broker

O Bandwidth Broker (BB) é um agente responsável pela alocação de serviços preferenciais para usuários no momento da requisição, e por configurar os roteadores da rede com o comportamento de entrega correto para o serviço definido. Um BB está associado a uma região de confiança particular, um por domínio; tem uma base de dados para policiamento que mantém as informações de quem pode fazer o quê, quando e um

método de utilizar a base de dados para autenticação de requisições. Apenas o BB pode configurar o roteador folha para entregar um serviço particular para um fluxo, crucial para o desenvolvimento de um sistema seguro.

Quando uma alocação é desejada para um fluxo particular, uma requisição é enviada para o BB. A requisição inclui o tipo de serviço, a taxa-destino, a rajada máxima e o período de tempo que o serviço será utilizado. A requisição pode ser realizada por um usuário ou ela pode vir de outras regiões de BB. Um BB autentica em primeiro as credenciais do requisitor, então, verifica se existe largura de banda disponível suficiente para a requisição. Se a requisição passa por este teste, a largura de banda disponível é reduzida pela quantidade requisitada e a especificação do fluxo é registrada.

O BB configura o roteador-folha com informações sobre o fluxo de pacotes a ser dado ao serviço no momento que o mesmo se inicia.

A idéia do BB foi introduzida como parte da arquitetura de Serviços Diferenciados. O BB está diretamente envolvido com a administração do gerenciamento de recursos de serviços diferenciados. Dois aspectos importantes são:

- Gerenciamento de recursos entre domínios;
- Gerenciamento de recursos dentro do domínio.

6.4.1.1 Gerenciamento de recursos dentro do domínio

O gerenciamento de recursos dentro do domínio refere-se à alocação de recursos dentro de uma rede ou domínio.

Em pequenas redes, os SLAs são negociados com clientes e as requisições de alocação de banda são aceitas ou rejeitadas dependendo da disponibilidade de recursos. E tanto informações sobre o DS code point a ser setado é passada ao roteador-folha, quanto informações sobre configuração de classes para o roteador de saída da rede.

Os Serviços Diferenciados são usados em redes de trânsito de modo a prover a usuários qualidade de serviço fim a fim, com a maior escalabilidade. Em redes de trânsito, o primeiro requisito é o gerenciamento de recursos para suportar fluxos de informações através do domínio, de um ponto de ingresso a um ponto de saída de rede. Quando um BB recebe uma mensagem de gerenciamento de recursos inter-domain, a mensagem contém: informações sobre um fluxo agregado entrante no domínio em um ponto de ingresso particular e um DS code point requisitado para o fluxo. Assim, o fluxo agregado existente, se houver um, entre o ponto de ingresso e de saída é atualizado com as informações do perfil de saída do fluxo. O BB observa o tráfego DS que entra e deixa o domínio, de modo a ter certeza que o acordo bilateral com os domínios adjacentes estão de acordo.

6.4.1.2 Gerenciamento de recursos entre domínios

O gerenciamento de recursos entre domínios refere-se ao provisionamento e alocação de recursos na borda da rede, entre dois domínios. Um Service-Level Agreement (SLA) bilateral, especificando a quantidade e os tipos de tráfego de cada lado, deve estar estabelecido nas bordas entre os dois domínios.

Mensagens de sinalização são enviadas entre BBs dos domínios adjacentes, de modo a requisitar os recursos necessários.

Em redes pequenas um bandwidth broker gerencia os recursos nos links conectados à redes de trânsito.

Em redes de trânsito, o BB observa o tráfego DS que entra e deixa o domínio, de modo a ter certeza que o acordo bilateral com o domínio adjacente esteja de acordo.

6.4.2 Arquitetura

O sistema Bandwidth Broker é formado dos seguintes componentes:

- A base de dados do Bandwidth Broker;
- O servidor Bandwidth Broker;
- A linha de comando do Bandwidth Broker;
- O cliente Service Level Agreement;
- A requisição do Bandwidth Allocation request;
- A configuração do roteador cliente do Bandwidth Broker.

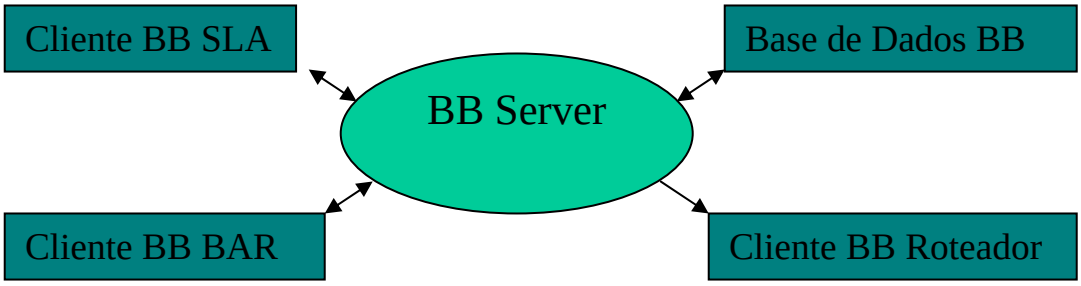

FIGURA 6.4.1 - A arquitetura Bandwidth Broker

6.4.3 A base de dados do Bandwidth Broker

Esta base de dados inclui um mecanismo para armazenagem de todos os dados relacionados à funcionalidade do Bandwidth Broker dentro da região de confiança.

Incluindo o seguinte:

- Service Level Agreements (SLA);
- Bandwidth Allocation Requests (BAR);
- Mapping of Bandwidth Allocation to Code Points.

Entradas são realizadas à base de dados sempre que uma requisição é aceita.

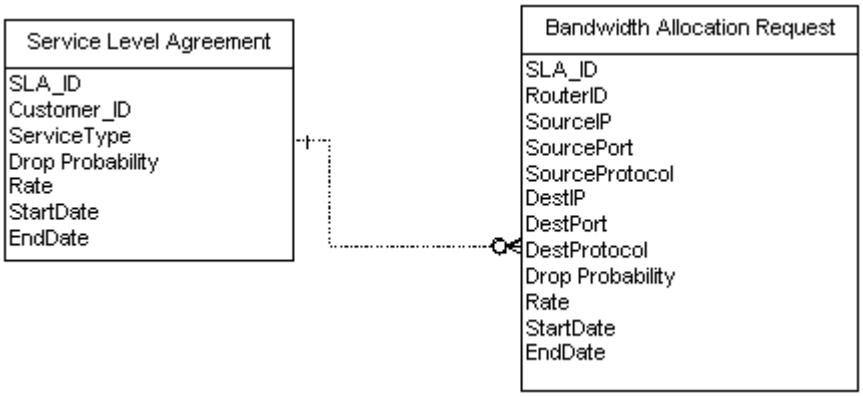

FIGURA 6.4.2 - A base de dados do BB

6.4.4 O servidor do Bandwidth Broker

O servidor do Bandwidth Broker é um programa multi-processos que manipula a interação entre a base de dados e as múltiplas requisições realizadas pelos clientes.

Ele realiza as seguintes funções:

- Realiza validação de dados e policia as atividades dos clientes (SLA or BAR);
- Se uma requisição é válida e dentro dos limites de alocação, mudanças correspondentes são feita na base de dados;
- Provê uma interface entre o cliente Bandwidth Broker e a base de dados do Bandwidth Broker;
- Provê informações de configuração para egress router de modo a configurar classes quando um novo SLA é aceito.
- Provê informações de configuração para o roteador folha quando um BAR é aceito, para que ele possa realizar policiamento e marcação do fluxo aceito.

6.4.5 A linha de comando do Bandwidth Broker

A linha de comando do Bandwidth Broker provê programas-clientes que interagem com o servidor Bandwidth Broker e a base de dados.

Dois tipos de comandos são encontrados:

- Service Level Agreements
- Requisições Bandwidth Allocation

6.4.5.1 Service Level Agreements:

O Service Level Agreements (SLA) provê um mecanismo simples para alocação de blocos de um serviço em particular para um cliente específico. Esta facilidade reserva banda para usuários ou organizações, mas não refere-se à alocação para fluxos específicos. Um SLA é válido para fluxo dentro de uma única região. Uma vez que o SLA tenha sido estabelecido, porções deste serviço podem ser associadas a fluxos específicos. O SLA inclui as seguintes informações:

- Identificação do cliente
- Tipo de serviço
- Parâmetros de tipo de serviço
- Restrições do serviço

O Bandwidth Broker assegura que todas as obrigações para qualquer serviço não exceda a quantidade deste serviço disponível na região de confiança.

6.4.5.2 Requisição de alocação de banda:

As requisições de alocação de banda são usadas por clientes de modo a requisitar porções de um serviço alocado por um SLA, para fluxos individuais.

Estes fluxos individuais são descritos por informações de origem/destino/protocolo, além da informação da taxa.

O pedido de requisição de banda contém:

- Identificador de usuário;
- O ID do SLA para negociar o SLA;
- Parâmetros de nível de serviço (taxa, rajada máxima, etc.);
- Identificador fonte (número da porta, endereço IP, protocolo);
- Destino (número da porta, endereço IP, protocolo);
- Duração da requisição.

O Bandwidth Broker, antes de permitir a requisição, garante que a alocação não irá violar os limites do SLA e também não excederá a quantidade do agregado do serviço na região de confiança.

6.4.6 Configuração do roteador

O Bandwidth Broker tem que configurar um grupo de roteadores com capacidades de DiffServ de modo a prover o nível desejado de serviço dentro da região.

O BB configura os roteadores-folhas e de saída no seu domínio, de acordo com o service level agreements e a requisição de alocação de banda de entrada para os clientes. Depois da verificação e validação de um SLA, o roteador de saída apropriado é conectado e parâmetros requeridos para a configuração do serviço particular são enviados. Similarmente, na validação de um BAR, o roteador-folha mencionado no BAR é conectado e os parâmetros necessários para marcação e policiamento do fluxo são passados a ele.

6.4.7 A comunicação com o cliente

Um cliente se comunica com o BB diretamente, enviando dados apropriados através da conexão da rede. A implementação usa o protocolo Bandwidth Broker Transfer Protocol (BBTP) como base para a interação.

O protocolo BBTP é baseado no paradigma request/response, no qual um cliente estabelece uma conexão como o servidor e envia uma mensagem de request para o servidor, consistindo da versão do protocolo, cabeçalho do request e a linha do request. O servidor responde com a mensagem da versão do protocolo, o status da linha, incluindo o código de sucesso ou falha, seguido pelo cabeçalho.

7 Avaliação do Diffserv

Os objetivos relacionados à evolução do Diffserv são similares aos requisitos e projetos para QoS na nova Internet. Do ponto de vista do QoS Working Group, o Diffserv deve ter ênfase na simplicidade, escalabilidade, interoperabilidade e administração. O que se segue é uma análise ponto a ponto de como o framework Diffserv está relacionado à qualidade de serviço requisitada pela nova Internet [TEI99c].

7.1 Habilita aplicações avançadas

De modo a desenvolver serviços que proporcionem os requisitos de QoS das aplicações avançadas, foram propostos quatro serviços com o critério de certamente garantir a transmissão fim a fim. São eles:

- Premium — emula uma linha privada. Peak bandwidth garantida com pouquíssimo atraso na fila, perda e jitter. Apropriado para aplicações intolerantes.
- Assegurado — emula uma rede quase sem carga. Similar a controle de carga; apropriado para aplicações tolerantes e adaptativas.
- Class of Service (CoS) — classe de serviço baseada em precedência. Melhor que o serviço tipo best-effort para encontrar as necessidades de priorização.
- Default — serviço do tipo best-effort.

O serviço Premium oferece um serviço com um limite de pico, muito pouca perda e jitter, e mínimo atraso na fila. Este serviço requer um PHB equivalente para observar prioridades nas filas, bem como políticas de controle de admissão que nunca vendem mais que a capacidade Premium de qualquer roteador. O serviço Premium é apropriado para a maioria das aplicações intolerantes e podem ser até implementadas pela comunidade de provedores, desenvolvedores e usuários, devido o seu contrato de serviço ser de fácil entendimento [TEI99c].

Serviços Assegurado é um exemplo de serviço previsível que ocupa a porção mediana do espectrum de serviços. Com Assegurado, os usuários contratam um perfil específico de serviço e é garantido que no perfil de tráfego existirá um uma rede sem carga, mesmo com a presença de congestionamento. Pacotes de acordo com o perfil são dados por um PHB que indica que eles devem ser descartados por último se houver a necessidade de descarte. Pacotes fora do perfil são remarcados por policiamento para o default – PHB do tipo best-effort. Serviços previsíveis como Assegurado são menos caros para provedores do que o tipo Premium, e podem alcançar as necessidades das aplicações que não requerem a garantia dada pelo tipo Premium. Tal serviço pode ser apropriado para aplicações que podem tolerar algum tipo de perda de pacote e apenas requerem transmissão média assegurada por um longo período de tempo.

Finalmente, o CoS será igualmente importante para os membros da nova Internet que desejam receber diferenciação em relação ao tráfego da rede ou para suportar aplicações que necessitam de serviços meramente melhores do que o empregado atualmente. Por exemplo, oferecer maior prioridade a e-mail em vez de chat para universidades. Entretanto, como o tráfego da Internet continua crescendo mais rápido que sua capacidade de provisionamento, o CoS nunca será uma saída para aplicações avançadas.

Segundo o QoS Working Group, é recomendado um testbed das formas de serviços oferecido pelo modelo Diffserv – especialmente os serviços Premium e Assegurado descritos acima. Adicionalmente, é recomendado que as redes já suportem funcionalidades de CoS para oferecer este serviço comercialmente.

7.2 Escalabilidade

Escalabilidade é um dos mais sérios requisitos de engenharia para QoS e esta é uma característica que a arquitetura Diffserv não tem motivo de queixas.

7.2.1 Escalabilidade para um grande número de fluxos

Roteadores do core são rotineiramente metralhados com milhares de reenvios de fluxos. Conseqüentemente, qualquer modelo de QoS que necessite conhecer o status por fluxo ou causar overhead computacional não será escalar. O modelo DiffServ procura levar a complexidade por fluxo para a periferia da rede onde os fluxos são menores. O objetivo é prover serviços por fluxo seguros através de policiamento de fluxos no roteador-folha e servir apenas um pequeno número de agregados no core. Como foi discutido, agregados são caracterizados por apenas um pequeno número de per-hop forwarding behavior (PHBs) simples, os quais reduzem em muito a complexidade de reenvio por pacote nos roteadores do core [TEI99c].

7.2.2 Escalando para altas velocidades

Em soma à necessidade de manter os PHBs simples, os agregados poderiam ser manipulados em altas velocidades, ocorrendo assim um aumento na necessidade de manipulação de fluxos individuais em alta velocidade. Isto é especialmente verdade em ambientes onde crescimento na largura de banda disponível da rede cria uma percepção de que a banda está completamente lotada. Aumentos dramáticos na largura de banda disponível inspira e habilita novos tipos de aplicações que cada vez mais dependerão da existência de dutos enormes. Quando este duto torna-se eventualmente congestionado, a nova aplicação irá desejar QoS assegurado que possa escalar a velocidades mais altas. O modelo de serviços diferenciados de jogar a complexidade por fluxo o mais perto do cliente final representa a melhor saída para escalar QoS por fluxo para altas velocidades.

7.3 Administração

No framework de serviços diferenciados, cada nuvem é livre para configurar seu policiamento e procedimentos de administração, desde que o acordo bilateral realizado com a outra nuvem seja honrado. Sendo a política de alocação de recursos na maioria das vezes sensitiva, o bandwidth broker passa a manipular as requisições externas sem um extensivo compartilhamento das informações de policiamento.

Pelo fato de que políticas e procedimento variam muito entre domínios, a administração flexível do DiffServ passa a ser importante para nuvens que contêm hosts finais. O advento da produção de QoS inevitavelmente será acompanhado por um novo controle de acesso, de acordo com as apropriadas políticas locais. Dentro de um domínio, o bandwidth broker é o ponto central para prover autenticação, autorização e accounting (AAA) e também é o ponto onde as políticas são administradas [TEI99c].

Alguns campos dentro da nova Internet talvez tenham uma estrutura hierárquica descentralizada que será melhor servida por uma arquitetura distribuída de bandwidth broker, a qual é consistente com a arquitetura diffserv e pode também ter características de tolerância à falhas.

7.4 Medição

O modelo Diffserv permite que um contrato de perfil de um serviço QoS possa ser comparado com métricas de performance de IP bem definidas. Um usuário com ferramentas de medição estará apto a auditorar seu tráfego, bem como provedores poderão auditorar seu tráfego agregado.

Em soma, provedores de serviços de rede poderão se utilizar de ferramentas para planejamento de rede, engenharia e debugging. Por exemplo, medições poderão ajudar no descobrimento de ponto de congestionamento que necessitam ser atualizados ou gerenciados cuidadosamente pelo bandwidth broker.

7.5 Admitindo múltiplas e interoperáveis implementações

A interoperabilidade, baseada em padrões de equipamentos e suporte para serviços fim a fim, tem sido o objetivo principal desde o início do Diffserv. O IETF está atualmente focado em especificar padrões para PHBs para o código do cabeçalho do pacote que identifique-os; isto deveria permitir a interoperabilidade no nível mais baixo. A interoperabilidade entre nuvens seria alcançada por padrões de PHBs globais e acordos simples de nível de serviço que podem ser concentrados em forma de serviço fim a fim.

7.5.1 Interoperabilidade de Equipamentos

Embora muitos trabalhos do IETF Diffserv Working Group venham sendo especificados para poucos padrões de PHBs e procurando um acordo de como associar os bits do campo diffserv para manter compatibilidade com roteadores que usam o campo ToS, outros trabalhos serão necessários para definir padrões de protocolos para [TEI99c]:

- Que hosts sinalizem seus requisitos para seus bandwidth brokers locais;
- Autenticar usuários pelo Bandwidth brokers;
- Que bandwidth brokers possam sinalizar para outros bandwidth brokers para end-to-end call setup;
- Que bandwidth brokers possam sinalizar roteadores-folhas com os parâmetros de policiamento por fluxo.

7.5.2 Interoperabilidade de nuvens de rede

A necessidade de concatenar a administração separada e serviços de múltiplas nuvens de rede dentro de serviços fim a fim tem sido um dos motivadores da arquitetura Diffserv desde seu início. O framework Diffserv satisfaz estas necessidades definindo serviços apenas nas bordas da nuvem e permitindo flexibilidade na implementação de serviços dentro de cada nuvem. O bandwidth broker abstrai as diferenças de como as nuvens são internamente formadas e provê um ponto de encontro comum para negociação de controle de admissão. Dentro de uma nuvem Diffserv, engenheiros de rede têm uma flexibilidade considerável em escolher entre tecnologias competidoras, equipamentos de rede de vendedores e alternativas de provisionamento. Em particular, o Diffserv deveria ser implementado completamente no nível IP, ou em casos de nuvens IP over ATM, serem mapeados Diffservs PHBs para classes de serviços ATM [TEI99c].

O modelo de serviços diferenciados é particularmente bem formado para suportar as mudanças administrativas de serviços QoS através de nuvens de administração separadas. O framework resultante é simples, acordos bilaterais entre provedores de serviço que são fáceis de entender e auditoria. Em soma, ocultando o provisionamento interno e detalhes de engenharia atrás da abstração do bandwidth broker, provedores de serviço de rede estão eufóricos com sua privacidade e flexibilidade – as quais eles necessitam – para operar independentemente e de forma competitiva. A separação do gerenciamento de recursos dentro do domínio e o controle de admissão do differentiated forwarding têm sido comparado com a separação do forwarding (simples) do roteamento (complexo). Se a abstração paralela for igualmente aceita, ela poderia resultar no aumento de suporte sofisticado para políticas e eficientes técnicas de gerenciamento de recursos que ocorram independentemente dos mecanismos de forwarding.

7.6 Suporte de sistema operacional e Middleware

Uma arquitetura Diffserv pode ser desenvolvida inicialmente usando as atuais aplicações de QoS juntamente com uma simples requisição de e-mail. É esperado, entretanto, que rapidamente ocorra a necessidade de aplicações estarem aptas a sinalizar seus requisitos de QoS dinamicamente, com suporte de APIs apropriadas e middleware.

O mecanismo RSVP pode ser rapidamente disponível nos modernos sistemas operacionais e podem ser usados como o protocolo de sinalização entre hosts e seu bandwidth broker local. Uma arquitetura para incorporar RSVP ao Diffserv foi definida em um Internet Draf recente. [BER99a}

7.7 Desenvolvimento Incremental, início em 1998

Existe uma alegre coincidência na evolução paralela do Diffserv e da Internet2. Enquanto ocorre um movimento imediato na direção de serviços diferenciados, direcionado pelos provedores de serviço de rede comerciais pela necessidade de oferecer CoS, desenvolve-se um conjunto de novas funcionalidades que são essenciais para este salto significativo com o qual poderemos implementar os serviços necessários por fluxo pela Internet2. Condicionador de tráfego, classificador, policiador e parte das funcionalidades propostas pelo bandwidth broker, ou existem atualmente ou irão existir no próxima geração de roteadores.

8 Mecanismos para implementar QoS

Esta seção, tem como objetivo apresentar os mecanismos já desenvolvidos pelo fornecedor Cisco Systems, no que condiz à QoS. São eles:

- Classificação
- Gerenciamento de congestionamento
- Mecanismos para evitar congestionamento
- Policiamento e conformação
- Sinalização
- Mecanismos de eficiência de link

8.1 Classificação

A classificação utiliza um descritor de tráfego para categorizar um pacote dentro de um grupo específico, e para definir o pacote e marcá-lo como acessível para manipulação de QoS na rede. Usando classificação de pacotes, podemos particionar o tráfego da rede dentro de múltiplos níveis de prioridade ou classes de serviços. Quando os descritores de tráfego são usados para classificar tráfego, a origem concorda seguir os termos do contrato e a rede promete qualidade de serviço. O policiamento de tráfego, tal como a característica de limite de taxa do Committed Access Rate (CAR) e conformação de tráfego, além de Frame Relay Traffic Shaping (FRTS) e Generic Traffic Shaping (GTS), usam um descritor de tráfego de pacotes – classificação – para assegurar o contrato [CIS99].

A classificação de pacotes é primordial para técnicas de policiamento que selecionam pacotes que cruzam elementos de rede, ou uma interface particular para diferentes tipos de serviços QoS. Por exemplo, podemos utilizar classificação para marcar certos pacotes para IP Precedence e podemos identificar outros como um fluxo RSVP específico.

Os métodos antigos de classificação eram limitados ao conteúdo do cabeçalho do pacote. Os métodos atuais de marcação de pacote para classificação permitem configurar informações em cabeçalhos de nível 2, 3 ou 4, ou até configurar informações dentro do payload do pacote.

Esta seção explica o IP Precedence, dando uma breve descrição dos tipos de classificação de tráfego providos pelo mecanismo. São discutidas as seguintes características [CIS99]:

- Policy-Based Routing
- QoS Policy Propagation via Border Gateway Protocol
- Committed Access Rate

8.1.1 IP Precedence

O uso do IP Precedence permite especificar a classe de serviço (CoS) para um pacote. São usados os três bits precedentes do campo ToS no cabeçalho IPv4 para este propósito.

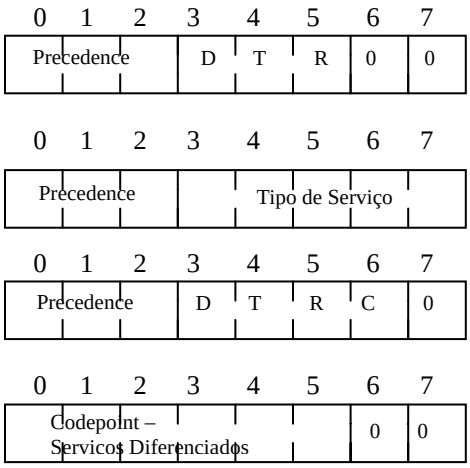

FIGURA 8.1.1 - Histórico do campo IP

Usando os bits ToS, podemos definir até seis classes de serviços. Outras características configuradas através da rede, podem então usar estes bits para determinar como tratar o pacote, em vez de considerar o tipo de serviço para garanti-lo. Estas outras características de QoS podem fornecer políticas apropriadas de manipulação de tráfego, incluindo estratégias de gerência de congestionamento e alocação de banda. Por exemplo, embora o IP Precedence não seja um método de enfileiramento, métodos de enfileiramento tais como: Weighted Fair Queueing (WFQ) e Weighted Random Early Detection (WRED) podem usar o IP Precedence para configurar os pacotes para tráfego priorizado [CIS99].

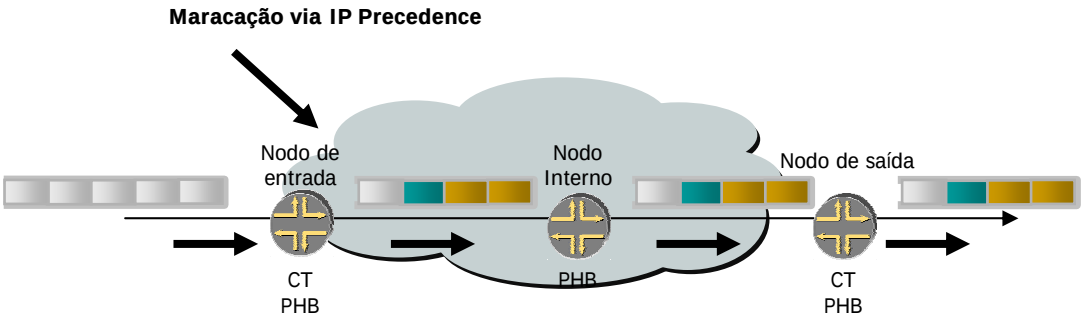

FIGURA 8.1.2 - IP Precedence

Configurando níveis de precedência em tráfego entrante e usando-os em combinação com características de enfileiramento, podemos criar serviços diferenciados. Igualmente pode-se utilizar características tal como policy-based-routing (PBR) e CAR para configurar a precedência, baseada em classificação por lista de acesso. Assim, cada elemento da rede pode prover serviços baseados em determinado policiamento, o IP Precedence é geralmente implementado o mais perto da borda da rede. Podemos pensar em IP Precedence com uma funcionalidade da borda que permite o core ou backbone, – características de QoS, tal como WRED–, reenviar tráfego baseado em CoS. O IP Precedence pode também ser configurado no host ou na rede do cliente, mas esta configuração pode ser reescrita por policiamento dentro da rede.

As seguintes técnicas podem utilizar o IP Precedence para determinar como o tráfego deve ser tratado:

- Distributed Weighted Random Early Detection (Distributed-WRED)
- Weighted Fair Queueing (WFQ)
- Committed Access Rate (CAR)

8.1.1.1 Como os bits do IP Precedence são utilizados para classificar pacotes

Podemos utilizar os três bits do campo ToS no cabeçalho IP para especificar o CoS dado para cada pacote, particionar o tráfego em até seis classes – os dois restantes são reservados para uso interno – e então usar mapas de policiamento e lista de acesso para definir políticas de rede, em termos de manipulação de congestionamento e alocação de bandwidth para cada classe.

Por razões históricas, cada precedente corresponde a um nome, os quais são definidos no RFC 791 (definição do protocolo IP) [RFC791]. A seguinte tabela lista os números e seus correspondentes nomes.

TABELA 8.1.1 - Valores do IP Precedence

Número	Nome
0	Rotineira
1	Prioritária
2	Imediata
3	Flash
4	Flash override
5	Crítica
6	Internet
7	Network

Embora características de IP Precedence permitam flexibilidade considerável para dar precedência, podemos definir nosso próprio mecanismo de classificação. Por exemplo, podemos querer dar precedência baseada em aplicações.

Nota: os bits IP Precedence configurados para 6 e 7 são reservados para informações de controle, tal como atualizações de roteamento.

8.1.1.2 Alterando os valores do IP Precedence

Por default, o valor do IP Precedence não é alterado, preservando-se o valor configurado no cabeçalho. Permiti-se assim que os dispositivos da rede possam prover serviços baseados no valor configurado [CIS99].

Esta política de policiamento segue o padrão, no qual o tráfego de rede deve ser agrupado dentro de vários tipos de serviços no perímetro da rede, e estes serviços devem ser implementados no core da rede. Os roteadores no core da rede podem utilizar os bits do IP Precedence, por exemplo, para determinar a ordem de transmissão, o seu descarte, e assim por diante.

Podemos utilizar um dos seguintes mecanismos para configurar IP Precedence nos pacotes:

- Policy-Based Routing
- QoS Policy Propagation via Border Gateway Protocol (PB-BGP)
- Committed Access Rate (CAR)

Como mencionado anteriormente, depois que um pacote tenha sido classificado, podemos usar outro mecanismo tal como CAR e WRED, para especificar e forçar policiamento para assegurar o modelo.

8.1.2 Policy-Based Routing

O PBR permite configurar IP Precedence, para especificar o caminho correto do tráfego, ou o caminho baseado em configurações de policiamento.

O PBR permite realizar o seguinte:

- Classificar o tráfego baseado em critérios de lista de acesso, assim estabelecendo critérios de associação;
- Configurar IP Precedence, dando à rede a habilidade de habilitar diferentes classes de serviços;
- Rotear pacotes para especificar caminhos; rotear para permitir serviços QoS através da rede.

O policiamento pode ser baseado em endereço IP, número da porta, protocolo, ou tamanho do pacote. Para um simples policiamento, podemos utilizar um destes descritores; para um policiamento complexo, podemos usar todos eles.

Por exemplo, a classificação de tráfego por PBR permite identificar tráfego por diferentes tipos de serviço na borda da rede e então implementar QoS definido por cada CoS no core da rede, usando técnicas de prioridade ou weighted fair queueing. Este processo obviamente necessita classificação de tráfego, explicitamente em cada interface no core da rede.

8.1.2.1 Como funciona

Todos os pacotes recebidos, com PBR habilitado na interface, são passados através de filtros conhecidos como mapas de rotas. O mapa de rotas usado pelo PBR dita o policiamento, determinando para onde os pacotes devem ser encaminhados.

8.1.2.2 Quando devemos utilizar Policy-Based Routing ?

Algumas aplicações ou tráfegos podem ser beneficiados por roteamento; por exemplo, poderíamos transferir registros de estoque para um escritório com alta largura de banda, enquanto transmitimos aplicações rotineiras tal como e-mail, através de links de baixa velocidade [CIS99].

8.1.3 QoS Policy Propagation via Border Gateway Protocol

O BGP é um protocolo de roteamento entre domínios, que troca informações de roteamento com outros sistemas BGP [RFC 1163].

Políticas de propagação via BGP permite classificar pacotes baseados no seguinte:

- Lista de acesso
- Lista de community BGP
- Caminhos de sistemas autônomos BGP
- IP Precedence
- Endereços de origem e destino

Depois do pacote ter sido classificado usando BGP, podemos usar características de QoS tal como CAR e WRED para especificar policiamento, de modo a atender o modelo do negócio.

8.1.4 Committed Access Rate

O CAR é um mecanismo que implementa classificação de serviços e policiamento através de limites de taxa.

Podemos utilizar serviços de classificação por CAR para configurar IP Precedence em pacotes que entram na rede. Esta característica do CAR permite particionar a rede em múltiplos níveis de prioridade ou classes de serviço. Dispositivos de rede dentro da rede podem utilizar o IP Precedence para determinar como o tráfego deve ser tratado.

Depois do pacote ser classificado, a rede pode aceitar ou reescrever e reclassificar o pacote, de acordo com um policiamento específico.

Assim, o Committed Access Rate (CAR) tem como objetivo prover ao operador da rede a característica de gerenciamento da alocação da largura de banda, a qual foi determinada na criação da conexão. O tratamento ocorre através de Token bucket e thresholds, o último podendo ocorrer, tanto em filas de entrada como em filas de saída dos roteadores. Entre as ações tomadas, pode ocorrer a mudança de classe (IP precedence) ou o descarte de pacotes (RED-like). Como exemplo de CAR Policy podemos citar:

- Firm CAR – pacotes que excedem a banda alocada são descartados;
- CAR + Premium – pacotes que excedem a banda alocada são “re-coloridos” com alta ou baixa preferência;
- CAR + Best Effort – pacotes que excedem a banda alocada são “re-coloridos” até o estouro do threshold, depois são descartados;
- Per Application CAR – diferentes CARs são especificados para diferentes aplicações.

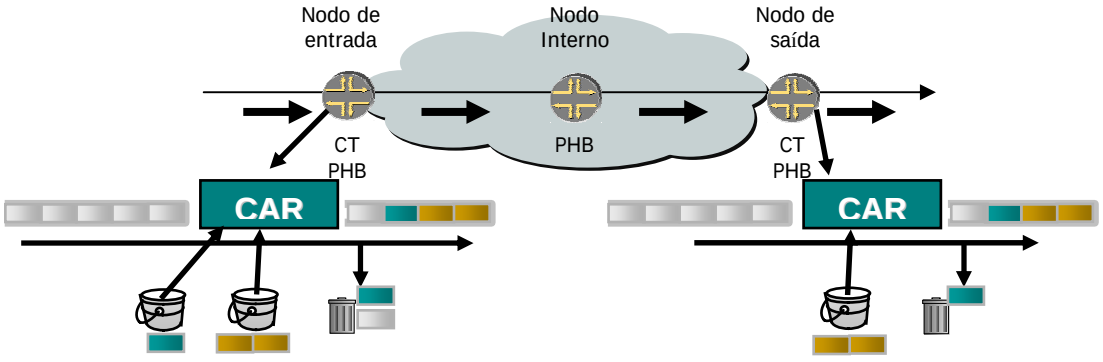

FIGURA 8.1.3 - Committed Access Rate

8.2 Gerenciamento de Congestionamento

As características de gerenciamento de congestionamento permitem o controle de congestionamento, pela determinação da ordem com que os pacotes são transmitidos para fora da interface, baseada em prioridades para estes pacotes. O gerenciamento de congestionamento está diretamente ligado à criação de filas; o direcionamento dos pacotes para estas filas baseia-se na classificação dos pacotes, e na programação dos pacotes em uma fila para transmissão. A característica de gerenciamento de congestionamento oferece quatro tipos de protocolos, cada um dos quais permite que se especifique a criação de um número diferente de fila, de acordo com o tipo de tráfego e a ordem na qual os pacotes são transmitidos [CIS99a].

Durante períodos onde não existe congestionamento, os pacotes são transmitidos para fora da interface assim que eles chegam. Durante períodos de congestionamento, os pacotes chegam mais rápido que a interface de saída pode suportar. Se for utilizado o gerenciamento de congestionamento, os pacotes acumulados na interface são enfileirados até que a interface esteja livre novamente; o envio dos pacotes é então programado para transmissão, de acordo com sua prioridade, e o mecanismo de enfileiramento configurado para a interface. O roteador determina a ordem com que os pacotes são transmitidos, controlando quais pacotes são colocados nas filas e como filas são servidas com respeito as outras.

Esta seção descreve estes quatro tipos de enfileiramento, os quais constituem os mecanismos de gerenciamento de congestionamento para QoS:

- First-In, First-Out Queueing (FIFO) – o método FIFO não utiliza o conceito de priorização ou classes de tráfego. Com FIFO, a transmissão de pacotes para fora da interface ocorre de acordo com a chegada dos mesmos;
- Weighted Fair Queueing (WFQ) – o WFQ divide o bandwidth através de filas de tráfego baseado em pesos. O WFQ assegura que todo o tráfego é tratado de acordo com as regras, dado seu peso. Para ajudar a entender como o WFQ trabalha, considere uma fila para pacotes FTP como uma fila coletiva, e uma fila para tráfego de pacotes interativos como uma fila individual. Dado o peso das filas, o WFQ assegura que para todos os pacotes da fila coletiva transmitidos, um número igual de pacotes da fila individual é transmitido. Assim, o WFQ assegura, satisfatoriamente, o tempo de resposta a aplicações críticas, tal como as interativas, aplicações baseadas em transações que são intolerantes a degradações de performance;
- Custom Queueing (CQ) – com CQ, o bandwidth é alocado proporcionalmente para cada classe de tráfego diferente. O CQ permite especificar o número de bytes ou pacotes para a fila. Neste caso, o CQ é utilizado geralmente para interface de baixa velocidade;
- Priority Queueing (PQ) – com PQ, pacotes com prioridade maior são enviados antes que todos os pacotes com prioridade menor, de modo a assegurar o tempo de entrega destes pacotes.

8.2.1 Por que usar Gerenciamento de Congestionamento?

Atualmente, existe uma necessidade real de que o tráfego seja compartilhado entre aplicações, de modo que não venha a afetar a performance das mesmas. Neste caso, devemos considerar cada vez mais a utilização de técnicas de gerenciamento de congestionamento para assegurar o tratamento através dos vários tipos de tráfego [CIS99a].

Aqui estão algumas idéias a serem consideradas sobre a necessidade real de se utilizar gerenciamento de congestionamento:

- A priorização de tráfego é especialmente importante para aplicações sensíveis ao atraso e transações interativas – por exemplo, vídeo- conferência – que necessitam prioridade maior que aplicações de transferência de arquivos. Entretanto, usar WFQ assegura que todo o tráfego é tratado de acordo com o seu peso. Por exemplo, o WFQ endereça os requisitos de aplicações interativas sem penalizar aplicações FTP;
- A priorização é mais efetiva em links WAN nos quais a combinação entre tráfego em rajadas e taxas menores de dados podem causar congestionamentos temporários;
- Dependendo do tamanho médio dos pacotes, a priorização é mais efetiva quando aplicada a links T1/E1 ou menores.
- Se os usuários de aplicações que rodam através da rede identificam uma resposta pobre em relação ao tempo, deve ser considerada a utilização de características de gerenciamento de congestionamento. Características de gerenciamento de congestionamento são dinâmicas, podendo se ajustar sozinha às condições existentes na rede. Entretanto, considerando que se um link WAN esta constantemente congestionado, a priorização de tráfego pode não resolver o problema. A melhor solução seria aumentar o tamanho do link.
- Se não existe congestionamento no link WAN, não há razão para implementar priorização de tráfego.

Aqui estão alguns passos que resumem os aspectos a serem considerados na determinação de quando devemos estabelecer e implementar um policiamento de fila para a rede:

Passo 1 – determine se o link WAN está congestionado, se algum usuário está percebendo alguma degradação de performance.

Passo 2 – determine seus objetivos baseado no tráfego a ser gerenciado, topologia e projeto de rede. Uma vez identificados os interesses, considere onde se encaixam os objetivos:

- Estabelecer um tratamento de distribuição de alocação de bandwidth para todos os tipos de tráfego identificados;
- Garantir um tratamento de priorização para tipos de aplicações especiais – por exemplo, aplicações multimídia interativas;
- Ajustar a alocação de bandwidth de modo que cada recurso de rede, compartilhado por toda a rede, possa possuir um determinado requisito de bandwidth identificado.

Passo 3 – configure a interface para o tipo de estratégia de enfileiramento escolhido e observe os resultados.

O padrão de tráfego altera todo o tempo, então devemos repetir o processo de análise do passo 2 periodicamente, e adaptar a configuração da fila de acordo com o resultado.

1.2.2 Decidindo qual policiamento de fila a ser usado

Esta seção destaca brevemente algumas diferenças entre os tipos de mecanismos de enfileiramento e inclui uma tabela que compara as três principais estratégias.

O mecanismo de FIFO não realiza nenhum tipo de priorização de pacotes no tráfego de dados do usuário. Ele não incorpora conceitos de priorização ou classe de tráfego. Quando o FIFO é usado, rajadas provenientes da fonte podem causar atrasos no tráfego sensível ao tempo ou no tráfego importante, e este pode ser descartado porque o tráfego de menor importância preencheu a fila. Considere as seguintes diferenças quando usar CQ ou PQ:

- o CQ garante alguns níveis de serviço para todo o tráfego, porque podemos alocar bandwidth para toda classe de tráfego. Podemos definir o tamanho da fila determinando sua capacidade de pacotes.
- o PQ garante priorização, assegurando que um tipo de tráfego seja transmitido, possivelmente antes que todos os outros. Para o PQ, uma fila de baixa prioridade pode ser muito afetada, e, no pior caso, permitir que nunca seja enviado pacotes.

TABELA 8.2.1 - Comparação entre as filas

	WFQ	**CQ**	**PQ**
Número de filas	• 256 por default	• 16 filas de usuários	• 4 filas
Tipos de serviços	• Diferencia os fluxos através de pesos	• Round robin service • Alocação proporcional de banda para diferentes classes de serviço	• Filas de maior prioridade são tratadas primeiro
Configuração	• Não requer	• Requer configuração	• Requer configuração

8.2.2 First-In, First-Out

Na sua simples forma, o enfileiramento FIFO – também conhecido como first-come, first-served (FCFS) – envolve armazenamento de pacotes quando a rede está congestionada e os retransmite em ordem de chegada quando a rede não está mais congestionada.

O mecanismo FIFO não engloba o conceito de prioridade ou classe de tráfego e conseqüentemente não faz decisões sobre priorização de pacotes. Existe apenas uma fila, e todos os pacotes são tratados da mesma maneira. Os pacotes são enviados pela interface, na

ordem em que eles chegaram. Pacotes de alta prioridade não são transmitidos antes dos pacotes de baixa prioridade.

Quando o FIFO é usado, fontes de dados podem consumir toda a banda, rajadas provenientes de uma fonte podem causar atrasos no tráfego sensível ao tempo, e tráfegos importantes podem ser descartados porque o tráfego de menor importância completou a fila.

O mecanismo de FIFO, que é o método de enfileiramento mais rápido, é eficiente para links grandes que têm pouco atraso e congestionamento mínimo.

8.2.3 Weighted Fair Queueing

TABELA 8.2.2 - O WFQ

WFQ
• Baseados em fluxos • Pesos, quando os pacotes são classificados • FQ, quando não há classificação

O WFQ é um método de programação automática que provê tratamento de alocação de banda para todo o tráfego de rede. Ele aplica prioridade, ou pesos, para identificar tráfego, para classificar tráfego dentro de conversações, e determinar quanta banda cada conversação é relativamente permitida para as outras conversações. O WFQ é um algoritmo baseado em fluxo, que programa simultaneamente tráfego interativo do início da fila para reduzir o tempo de resposta, e trata o compartilhamento da banda restante entre fluxos de alta largura de banda. Em outras palavras, o WFQ permite dar a tráfegos de pouco volume, tal como sessões Telnet, maior prioridade sobre tráfego de alto volume, tal como sessões FTP. O WFQ oferece a transferência de arquivos concorrentes a flexibilidade de balanceamento da capacidade do link, ou seja, quando múltiplas transferências de arquivos ocorrem, para cada transferência é dada a mesma banda.

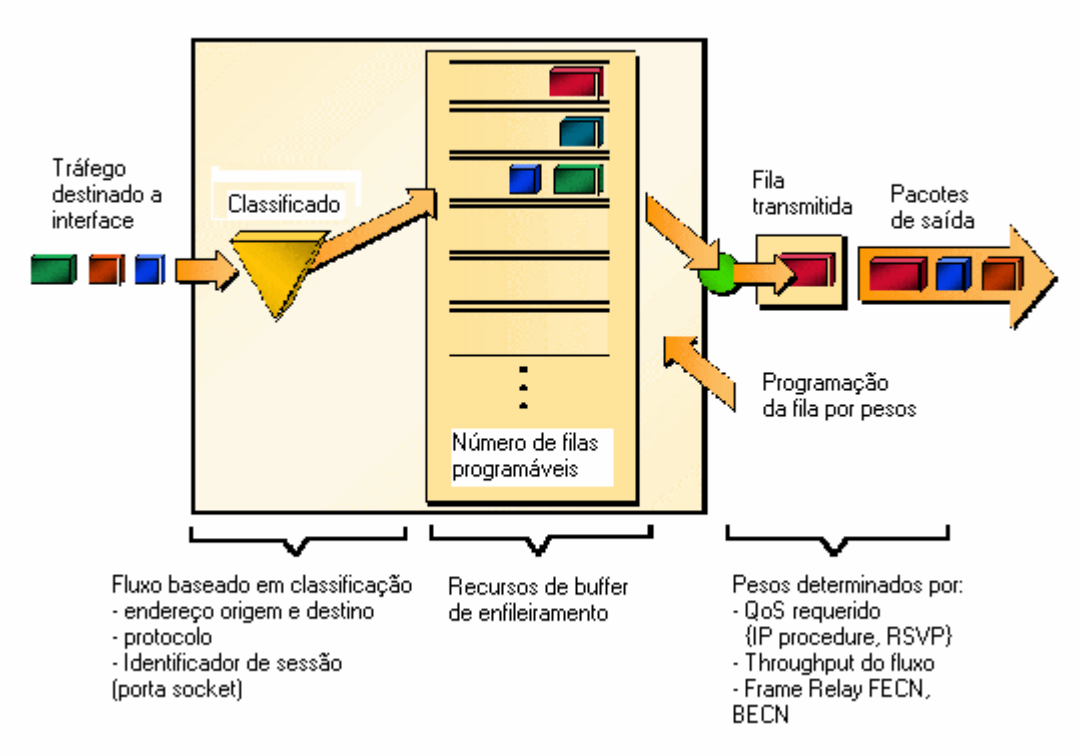

FIGURA 8.2.1 - Como o WFQ trabalha

O WFQ supera o mecanismo FIFO em vários aspectos. Quando o FIFO é implantado, o tráfego é transmitido na ordem que é recebido sem se importar com o consumo de banda ou atrasos associados. Como resultado, há transferência de arquivos e outras aplicações com alto volume de rede, a maioria das vezes degradando uma série de pacotes de dados associados. Estes pacotes relacionados são conhecidos como pacotes “trains”. Pacotes “trains” são grupos de pacotes que tendem a se movimentar juntos através da rede. Estes podem consumir toda a banda disponível, desprovendo outros tráfegos de utilizá-la.

O WFQ provê gerenciamento de priorização de tráfego que dinamicamente agrupa o tráfego dentro de mensagens que tornam-se conversações. O WFQ quebra os pacotes em partes dentro da conversação para assegurar que a banda seja compartilhada entre conversações individuais e que pequenos volumes de tráfego sejam transferidos em um determinado espaço de tempo [CIS99a].

O WFQ classifica o tráfego dentro de diferentes fluxos, baseados no endereço do cabeçalho do pacote, incluindo características tais como: endereço MAC ou de rede de destino e origem, protocolo, porta de origem e destino e número do socket da sessão, valores de DLCI (Frame-Relay data-link connection identifier), e valores de ToS (type of service). Existem duas categorias de fluxos: sessões com alta largura de banda e sessões com baixa largura de banda. Tráfego com baixa largura de banda tem prioridade efetiva sobre tráfego de alta largura de banda, e o tráfego de alta largura de banda compartilha o

serviço de transmissão proporcionalmente de acordo com os pesos assinalados. Fluxos de tráfego de baixa largura de banda, que são a maioria, recebem serviço preferencial, transmitindo toda a carga oferecida em um determinado espaço de tempo. No caso de fluxos de alta largura de banda, eles compartilham a capacidade restante proporcionalmente entre eles.

O WFQ coloca os pacotes de várias conversações na fila de tratamento, antes de transmiti-los. A ordem de remoção de uma fila de tratamento é determinada pelo tempo virtual de entrega do último bit de cada pacote que chega.

Novas mensagens para fluxos de alta largura de banda são descartados logo após o encontro da mensagem de congestionamento de threshold. Entretanto, fluxos de baixa largura de banda, os quais incluem mensagens de controle de conversação, continuam sendo enfileiradas. Como resultado, a fila de tratamento pode ocasionalmente conter mais mensagens que são especificadas pelo número de threshold.

O WFQ pode gerenciar fluxos de dados duplos, tal como entre pares de aplicações, e fluxos de dados simples tal como voz e vídeo.

O algoritmo WFQ também endereça o problema da variação de atraso ocorrido no round-trip. Se múltiplas conversações de alto volume estão ativas, suas taxas de transferência e períodos entre chegadas são muito mais previsíveis.

O WFQ provê uma solução para situações nas quais é desejado prover tempo de resposta consistente para usuários leves e pesados sem adicionar largura de banda. O WFQ se adapta automaticamente a mudanças nas condições de tráfego de rede.

8.2.3.1 Restrições

O WFQ não é suportado em soluções de tunelamento e encriptação, porque estas características modificam o conteúdo do pacote, o que é necessário para a classificação do WFQ.

8.2.3.2 O WFQ e o IP Precedence

O WFQ reconhece o IP Precedence. Ele detecta os pacotes com maior prioridade através da marcação de precedência, podendo programá-lo rapidamente, provendo tempo de resposta superior para este tráfego. Assim, à medida que a precedência aumenta, o WFQ aloca mais largura de banda para a conversação durante os períodos de congestionamento.

O WFQ dá um peso para cada fluxo, o qual determina a ordem de transmissão para os pacotes enfileirados. Neste tipo de esquema, pesos baixos são servidos primeiros.

Como o CQ, o WFQ transmite um certo números de bytes de cada fila. Com WFQ, cada fila responde a diferentes fluxos. Para cada ciclo através de todos os fluxos, o WFQ transmite efetivamente um número de bytes igual ao precedente do fluxo mais um.

Na verdade, este número é apenas usado como uma razão para determinar quantos bytes/pacotes serão transmitidos. Entretanto, de modo a entender o WFQ, o uso deste número como byte count é suficiente. Por exemplo, o tráfego com valor de campo IP Precedence igual a 7 possui um peso menor que um tráfego com valor de campo IP Precedence igual a 3.

Para determinar a banda alocada para cada fila, divide-se o byte count do fluxo pelo byte count total de todos os fluxos.

Por exemplo, no caso de termos um fluxo em cada nível de precedência, cada fluxo terá precedence +1 partes do link [CIS99a]:

1+2+3+4+5+6+7+8 = 36

Assim, a precedência 0 terá 1/36 da banda, a precedência 1 terá 2/36, a precedência 7 terá 8/38.

Entretanto, se tivermos 18 fluxos na precedência 1 e 1fluxo no restante, o total seria:

1+2(18)+3+4+5+6+7+8=70

A precedência 0 teria 1/70, e cada fluxo da precedência 1 teria 2/70.

Como fluxos são somados e diminuídos, a alocação de banda muda continuamente.

8.2.3.3 O WFQ e o Resource Reservation Protocol (RSVP)

O RSVP se utiliza do WFQ para alocar espaços em buffers e programar pacotes, e garante largura de banda para servir os fluxos. O WFQ trabalha com RSVP para ajudar a prover serviços garantidos e diferenciados de QoS. O RSVP é um padrão de protocolo Internet do IETF (RFC 2205) o qual permite a uma aplicação, dinamicamente, reservar banda de rede. O RSVP habilita a aplicações a requisição de um QoS específico para um fluxo de dados [CIS99a].

O RSVP é o único protocolo de sinalização padronizado para garantir largura de banda fim a fim em redes IP. Hosts e roteadores usam RSVP para entregar pedidos de QoS para roteadores ao longo de caminhos de fluxos de dados e para manter o status de roteadores e hosts para prover o serviço requisitado, geralmente banda e latência. O RSVP se utiliza da taxa de dados resultantes, da maior quantidade de dados que o roteador pode manter na fila, e do QoS mínimo para determinar a reserva de banda.

O WFQ ou WRED agem como o preparador para RSVP, configurando a classificação de pacotes e programando as requisições para os fluxos reservados. Usando o WFQ, o RSVP pode entregar um Integrated Services Guaranteed Service.

8.2.3.4 O WFQ e o Frame Relay

Os pesos de WFQ são afetados pelos bits ***discard eligible*** (DE), ***forward explicit congestion notification*** (FECN), e ***backward explicit congestion notification*** (BECN), do Frame Relay, quando o tráfego é comutado pelo Frame Relay. Uma vez que o congestionamento é marcado, o peso usado pelo algoritmo é alterado, podendo não ser visualizado o congestionamento na outra borda.

8.2.3.5 Considerações

Embora o WFQ se adapte automaticamente a mudanças nas condições do tráfego de rede, ele não oferece o controle preciso sobre alocação de banda que o CQ faz.

8.2.4 Custom Queueing

O CQ permite que seja especificado um número de bytes a serem repassados de uma fila, cada vez que a fila é servida, ou seja, permite que sejam compartilhados recursos entre aplicações com uma banda mínima específica ou requisitos de latência. Podendo, também, especificar um número máximo de pacotes em cada fila.

8.2.4.1 Como trabalha

O CQ manipula o tráfego, especificando o número de pacotes ou bytes a serem servidos por cada classe de tráfego. Ele serve às filas realizando ciclos – através delas – do tipo round-robin, enviando a porção da banda alocada para cada fila, antes de mover-se para a próxima fila. Se um fila estiver vazia, o roteador enviará pacotes da próxima fila que tiver pacotes prontos para serem enviados.

Quando o CQ é habilitado na interface, o sistema mantêm 17 filas de saída para aquela interface. Podendo especificar filas de 1 a 16. Associado à cada fila de saída está uma configuração de byte count, a qual especifica quantos bytes de dados o sistema deverá entregar da fila corrente, antes de mover-se para a próxima fila.

A fila de número 0 é uma fila de sistema; ela é esvaziada antes que as filas de 1 a 16 sejam processadas. A fila de sistema possui os pacotes de mais alta prioridade, tais como: pacotes keepalive e pacotes de sinalização. Nenhum outro tráfego pode ser configurado para ser usado nesta fila.

Para as filas de número 1 a 16, o ciclo entre as filas é seqüencial (round-robin fashion), desenfileirando o byte count configurado de cada fila em cada ciclo, entregando pacotes da fila corrente antes de se mover para a próxima. Quando uma fila em particular começa a ser processada, os pacotes são enviados até que o número de bytes enviados exceda o byte count da fila ou até que a fila esteja vazia.

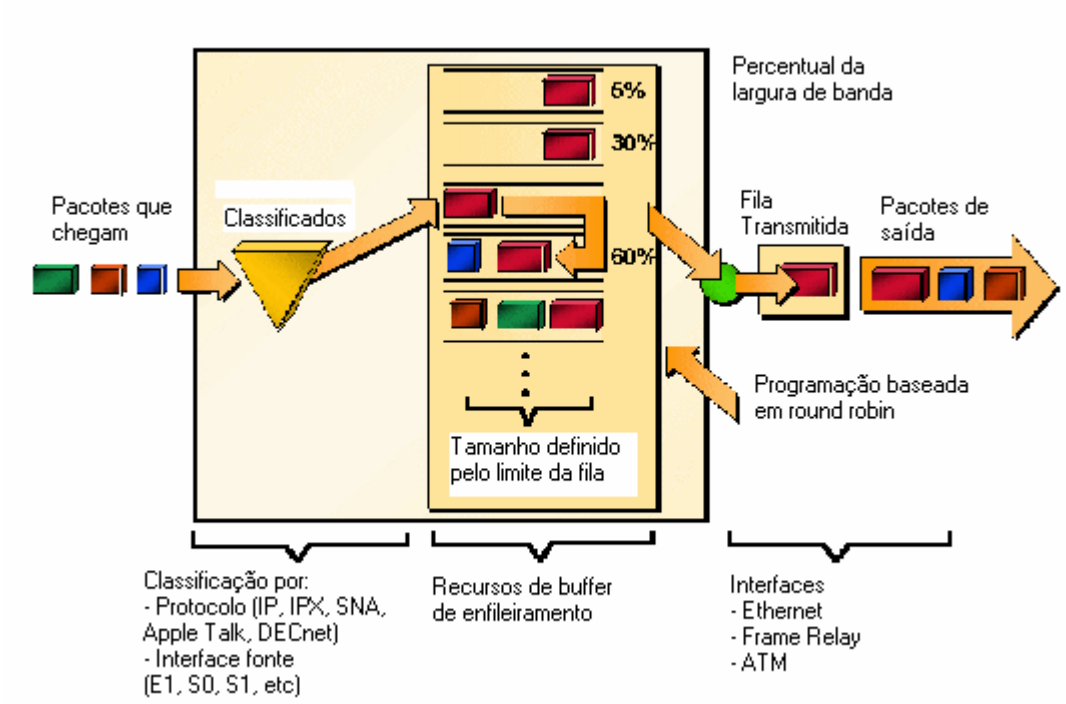

FIGURA 8.2.2 - Comportamento do CQ

O CQ assegura que nenhuma aplicação ou grupo específico de aplicações alcancem mais que uma porção pré-determinada de toda a capacidade, quando a linha está sobrecarregada. Como PQ, o CQ é estaticamente configurado e não se adapta automaticamente a condições de mudanças de rede.

8.2.4.2 Determinando os valores do Byte Count para a fila

De modo a alocar largura de banda para diferentes filas, podemos identificar o byte count para cada fila.

8.2.4.3 Como o byte count é usado

O roteador envia pacotes de um fila em particular até que o byte count exceda. Uma vez que o valor do byte count excedeu, o pacote que está atualmente sendo transmitido será

completamente enviado. Deste modo, se setarmos o byte count para 100 bytes e o tamanho do pacotes é 1024 bytes, então toda vez que esta fila é servida, 1024 bytes serão enviados, não 100 bytes.

Por exemplo, suponhamos que um protocolo tenha 500 bytes de pacotes, outro tenha 200 bytes de pacotes, e o terceiro tenha 100 bytes de pacote. Se desejarmos dividir a banda entre os três protocolos, podemos especificar um byte count de 200, 200 e 200 para cada fila. Entretanto, esta configuração não resulta em uma razão de 33/33/33. Quando o roteador serve a primeira fila, ele envia um único pacote; quando ele serve a segunda fila, ele envia 300 bytes; e quando ele serve a terceira fila, ele envia dois pacotes de 100 bytes. A razão efetiva no caso é 50/30/20.

Assim, configurar o byte count muito pequeno pode resultar em uma alocação de banda não desejada.

Entretanto, um byte count muito grande pode ocasionar outros problemas. Se endereçarmos 10 KB, 10 KB e 10 KB para as três filas do exemplo acima, cada protocolo, ao ser servido, levará muito tempo para ser servidor novamente. A melhor solução seria especificar 500 bytes, 600 bytes e 500 bytes de count para cada fila respectivamente. Esta configuração resulta em uma razão de 31/38/31, a qual pode ser aceitável.

De modo a configurar o byte count das filas da melhor maneira possível, devemos determiná-lo baseado no tamanho do pacote de cada protocolo.

8.2.4.4 Determinando o Byte Count

Para determinar o byte count correto, realiza-se as seguintes tarefas:

Passo 1 – para cada fila, divide-se a porcentagem de banda que desejamos alocar para a fila, pelo tamanho do pacote, em bytes. Por exemplo, assuma que o tamanho do pacote para o protocolo A é 1086 bytes, o protocolo B é 291 bytes, e o protocolo C é 831 bytes. Desejamos alocar 20% da banda para A, 60% para B e 20% para C. A razão deveria ser: 20/1086, 60/291, 20/831 ou 0.01842, 0.20619, 0.02407.

Passo 2 – Normalizam-se os números através da divisão pelo número menor:1, 11.2, 1.3. O resultado é a razão do número de pacotes que devem ser enviados. Então a porcentagem de banda que cada protocolo usa será de aproximadamente 20%, 60% e 20%.

Passo 3 – A fração, em qualquer um dos valores das razões, significa que pacotes adicionais serão enviados. Neste exemplo, a razão atual seria 1 pacote, 12 pacotes e 2 pacotes.

Passo 4 – Converte-se a razão do número de pacotes e byte count através da multiplicação de cada packet count pelo tamanho do pacote correspondente. Neste exemplo, o número de pacotes enviados será de um pacote com 1086 bytes, 12 pacotes com 291 bytes e dois pacotes com 831 bytes ou 1086, 3492 e 1662 bytes, respectivamente, de

cada fila. Estes seriam os byte counts que deveriam ser especificados na configuração da fila.

Passo 5 – Para determinar a distribuição de banda que esta razão representa, primeiro determine o número total de bytes enviados de todas as três filas: (1 x 1086) + (12 x 291) + (2 x 831) = 1086 + 3492 + 1662 = 6240.

Passo 6 – Então determina-se a porcentagem do número total de bytes enviados de cada fila: 1086/6240, 3492/6240, 1662/6240 = 17.4%, 56%, e 26.6%. Como podemos ver, os valores são muito próximos da razão desejada de 20/60/20.

Passo 7 – Se a banda atual não está próxima o suficiente à banda desejada, multiplica-se a razão original de 1:11.2:1.3 pelo melhor valor, tentando se aproximar o mais perto dos valores inteiros. Note-se que o multiplicador usado não necessita ser um inteiro. Por exemplo, se multiplicamos a razão por dois, conseguimos 2:22.4:2.6. Então, enviaremos dois pacotes de 1086 bytes, 22 pacotes de 291 bytes e três pacotes de 831 bytes, ou 2172/6693/2493, para um total de 11.358 bytes. A razão resultante é 19%/59%/22%, a qual é muito próxima do desejado.

8.2.4.5 Tamanho da Janela

O tamanho da janela também afeta a distribuição da banda. Se o tamanho da janela de um protocolo particular é configurado para um, então, o protocolo não colocará nenhum outro pacote dentro da fila, até que ele receba um reconhecimento. O algoritmo do CQ move-se para a próxima fila, se o byte count excedeu ou se não há pacotes na fila.

Deste modo, com o tamanho da fila em um, apenas um frame será enviado por vez. Se o frame count é configurado em 2 KB, e o tamanho do frame é de 256 bytes, então, apenas 256 bytes serão enviados por cada vez que a fila é servida.

8.2.4.6 Por que usar Custom Queueing?

O CQ pode ser usado para prover banda garantida para um tráfego específico em um ponto de congestionamento em potencial, assegurando ao tráfego uma porção fixa de banda disponível e deixando o restante da banda para outros tráfegos. Por exemplo, poderíamos reservar metade da banda para dados SNA, permitindo que a outra metade possa ser usada por outros protocolos. Se um tipo particular de tráfego não utilizar a banda reservada para ele, então, a banda não utilizada pode ser alocada dinamicamente para outros tipos de tráfego [CIS99d].

8.2.4.7 Considerações

O CQ é configurado estaticamente e não se adapta às condições de mudanças da rede. Com o CQ habilitado, o sistema leva mais tempo para comutar pacotes que FIFO porque os pacotes são classificados.

8.2.5 Priority Queueing

O PQ permite definir como o tráfego é priorizado na rede. Podemos configurar quatro prioridades de tráfego, e definir uma série de filtros nos pacotes, para fazer com que o roteador coloque o tráfego dentro das quatro filas; a fila com a mais alta prioridade é servida primeiro, até que seja esvaziada, então, as filas de menor prioridade são servidas em seqüência.

8.2.5.1 Como trabalha

Durante a transmissão, o PQ dá ao tráfego importante a mais alta prioridade, sempre levando precedência sobre o tráfego de menor importância. Os pacotes são classificados segundo os critérios especificados pelo usuário, e colocados dentro de uma das quatro filas de saída – alta, média, normal e baixa – baseada na priorização. Os pacotes não classificados caem dentro da fila normal.

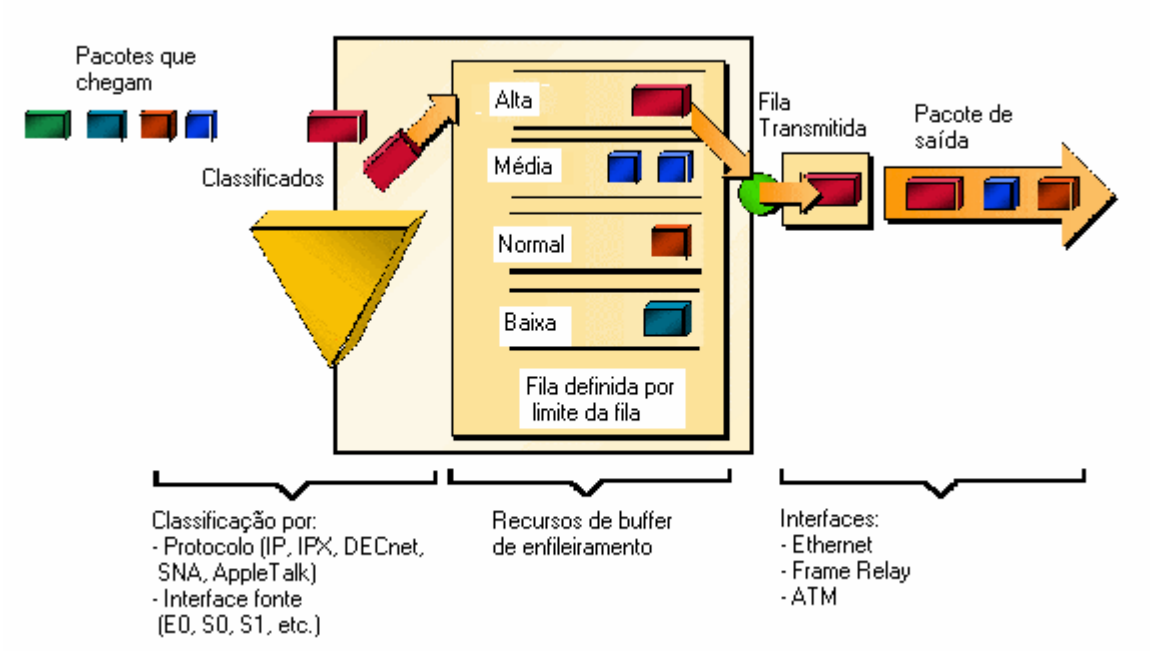

FIGURA 8.2.3 - Mecanismo Priority Queueing

Quando um pacote está para ser enviado para fora da interface, as filas de prioridades na interface são escaneadas por pacotes, na ordem descendente de prioridade. A fila de prioridade mais alta é escaneada primeiro, então, a fila de prioridade média, e assim por diante. O primeiro pacote que se encontra na fila de maior prioridade é escolhido primeiro para transmissão. Este procedimento é repetido toda vez que um pacote está por ser transmitido.

O tamanho máximo de uma fila é definido pelo tamanho limite. Quando uma fila encontra-se maior que o limite da fila, todos os pacotes a mais são descartados.

8.2.5.2 Como os pacotes são classificados para Priority Queueing

Uma lista de prioridades é um conjunto de regras que descrevem como os pacotes devem ser assinalados para filas de prioridades. Um lista de prioridade pode também descrever uma prioridade default ou o limite do tamanho da fila de várias filas de prioridade.

Os pacotes podem ser classificados pelo seguinte:

- Tipo de protocolo ou subprotocolo
- Interface de entrada
- Tamanho do pacote
- Fragmentos
- Lista de acesso

Pacotes do tipo keepalive são sempre assinalados como pacotes de alta prioridade; todos os outros tráfegos de gerenciamento (tal como IGRP) devem ser configurados. Os pacotes que não são classificados pelo mecanismo de priorização são assinalados como prioridade normal.

8.2.5.3 1.2.6.3 Por que usar Priority Queueing?

O PQ provê tratamento preferencial absoluto para tráfego de prioridade alta, assegurando que o tráfego de missão crítica atravesse vários links de WAN, tendo tratamento de prioridade. Em soma, o PQ provê uma resposta mais rápida que outros métodos de enfileiramento.

Embora possa ser utilizado em qualquer tipo de interface, a sua melhor utilização ocorre em interfaces de baixa largura de banda.

8.2.5.4 Considerações

Na escolha do PQ, considere-se que na maioria das vezes o tráfego de mais baixa prioridade não recebe banda para transmitir devido ao fato do tráfego de maior prioridade ter preferência. O uso de PQ, no pior caso, pode resultar em que o tráfego de mais baixa prioridade nunca seja transmitido. De modo a evitar esta condição em tráfego de baixa prioridade, podemos nos utilizar de conformação de tráfego ou CAR para limitar a taxa do tráfego de mais alta prioridade.

O PQ introduz um overhead extra que é aceito por interfaces lentas, mas pode não ser aceito por interfaces de mais alta velocidade, tal como Ethernet. Com o PQ habilitado, o sistema leva um tempo maior para comutar os pacotes, uma vez que os mesmos são classificados.

O PQ usa configuração estática e não se adapta às condições de mudanças da rede.
Restrições: o PQ não é suportado em tunelamento.

8.3 Mecanismos para evitar congestionamento

As técnicas para evitar congestionamento monitoram a carga do tráfego de rede, de modo a antecipar e evitar o congestionamento em épocas de gargalos de rede. Evitar congestionamento tem como base o descarte de pacotes. Entre as mais variadas técnicas de evitar congestionamento usadas, encontramos a Random Early Detection (RED), a qual se destina a redes de transmissão de alta velocidade [CIS99a].

Esta seção disponibiliza uma breve descrição dos tipos de características para evitar congestionamento. Entre elas temos:

Tail Drop – Esta técnica é a padrão para evitar comportamentos de congestionamento.

Weighted Random Early Detection (WRED) – Combina as características do algoritmo RED com IP Precedence.

8.3.1 Tail Drop

O mecanismo de tail drop trata todo tráfego da mesma maneira e não faz diferenciação entre as classes de serviços. As filas são preenchidas em períodos de congestionamento. Quando a fila de saída é completada e o mecanismo de tail drop está em vigor, os pacotes são descartados até que o congestionamento seja eliminado e a fila não esteja mais cheia.

8.3.2 Weighted Random Early Detection

Esta seção oferece uma introdução breve dos conceitos de RED e endereça o WRED, uma implementação de RED.

8.3.2.1 Random Early Detection

O mecanismo RED foi proposto por Sally Floyd e Van Jacobson em 1990, para endereçar congestionamento de rede em resposta à maneira tradicional. O mecanismo RED está baseado na premissa de que a maioria do tráfego roda em implementações de transporte de dados, as quais são sensíveis à perda, e em determinados períodos sofre um atraso devido ao descarte do seu tráfego. O TCP, que responde apropriadamente ao descarte do tráfego através de técnicas de atraso no envio do mesmo, permite a utilização do RED com um mecanismo de sinalização para evitar congestionamento.

É importante considerar que a utilização do RED deve ser empregada em transportes de rede tal como TCP, onde o protocolo é robusto, em resposta à perda de pacotes. No caso do protocolo Novell Netware e AppleTalk, nenhum deles é robusto em resposta à perda de pacotes, assim não devemos utilizar RED nestes casos.

8.3.2.2 Como trabalha

O objetivo do RED é controlar o tamanho médio da fila indicando aos hosts quando eles devem transmitir seus pacotes mais lentamente.

O RED leva vantagem ao utilizar-se do mecanismo de controle de congestionamento do TCP. Através do descarte randômico de pacotes em períodos de grande congestionamento, o RED conta a origem dos pacotes nos quais deve ocorrer uma diminuição na sua taxa de transmissão. Assumindo que o pacote de origem está utilizando TCP, a fonte irá diminuir sua taxa de transmissão até que todos os pacotes possam alcançar o seu destino, indicando que o congestionamento não ocorre mais. Na verdade, o TCP não pára totalmente, ele restarta rapidamente e adapta-se à taxa de transmissão que a rede pode suportar.

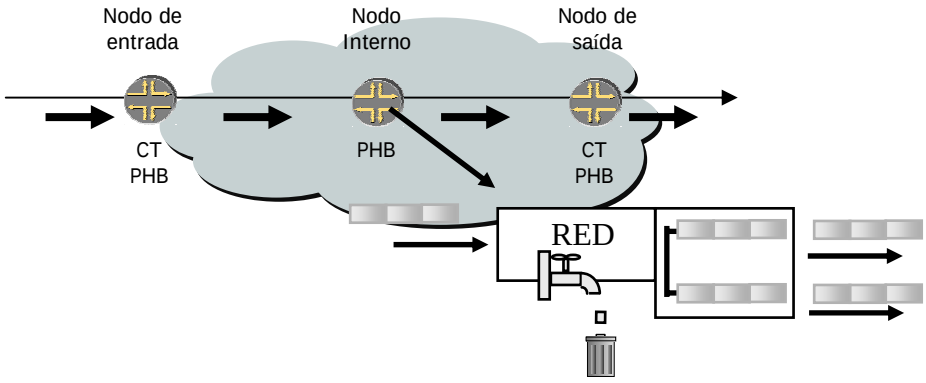

FIGURA 8.3.1 - Random Early Detection

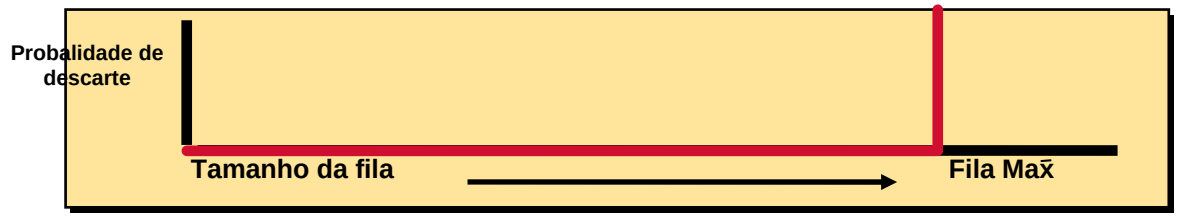

FIGURA 8.3.2 - Tráfego sem Random Early Detection

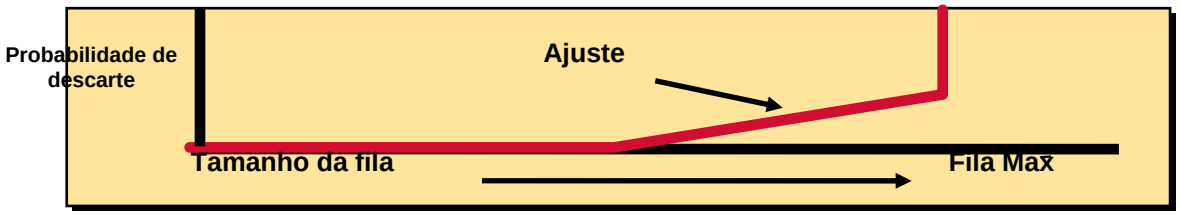

FIGURA 8.3.4 - Tráfego com Random Early Detection

- No primeiro gráfico, sem RED, nenhum pacote foi descartado, enquanto não chegou até o fim da fila. Porém, quando chegou ao fim da fila todos os pacotes foram descartados de uma vez só.
- O segundo gráfico mostra como o RED age. Quando não há pacotes na fila, o RED não tem efeito. Quando a fila alcança o threshold, o RED começa a descartar pacotes até baixar a taxa. Isso faz com que servidores TCP enviem mais devagar, aliviando o congestionamento. Se a fila continua a crescer, o RED começa a descartar mais agressivamente.

8.3.2.3 Probabilidade de Descarte de Pacotes

A probabilidade de descarte de pacotes é baseada em um minimum threshold, um maximum threshold e um mark probability denominator.

Quando a profundidade média da fila está acima do minimum threshold, o RED inicia o descarte dos pacotes. A taxa de descarte dos pacotes aumenta linearmente à medida que o tamanho médio da fila aumenta, até que o tamanho médio da fila alcance o maximum threshold.

O mark probability denominator é a fração de pacotes descartados quando a profundidade média da fila está no maximum threshold. Por exemplo, se o denominator é 512, um para cada 512 pacotes é descartado quando a média da fila está no maximum threshold.

Quando o tamanho médio da fila é maior que o maximum threshold, todos os pacotes são descartados. A seguinte figura mostra a probabilidade de descarte de um pacote.

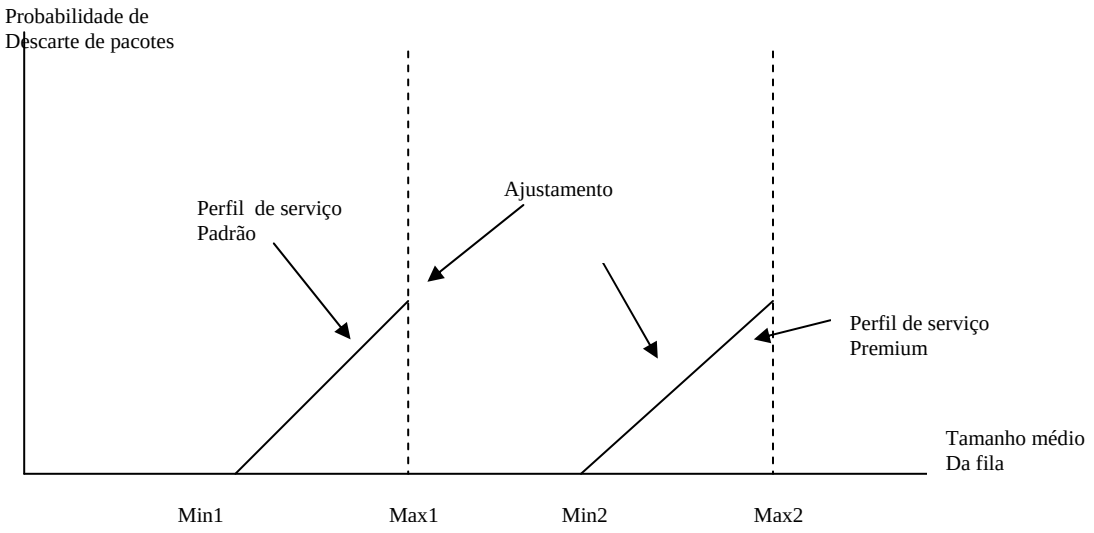

FIGURA 8.3.5 - Probabilidade de descarte no RED

O valor de minimum threshold deveria ser configurado com grandeza suficiente para maximizar a utilização do link. Se o minimum threshold é muito baixo, os pacotes podem ser descartados desnecessariamente, e a transmissão do link não será usada totalmente.

A diferença entre o maximum threshold e o minimum threshold deveria ser grande o suficiente para evitar sincronização global. Se a diferença é muito pequena, muitos pacotes podem ser descartados de um vez, resultando em uma sincronização global [CIS99a].

8.3.2.4 Como o TCP manipula a perda de tráfego

Quando o recipiente do tráfego TCP – chamado de receptor – recebe o segmento de dados, ele verifica se os quatro octetos, os quais identificam o número de seqüência , estão de acordo com o esperado, indicando, assim, que o segmento de dados foi recebido em ordem. Se o número bate, o receptor envia todos os dados para a aplicação destino, para então atualizar o número de seqüência de modo a refletir o próximo número, e finalmente transmitir um reconhecimento (ACK) para o emissor ou programa – um ACK – para ser transmitido após um pequeno período de tempo. O ACK notifica o emissor que o receptor recebeu todos os segmentos [CIS99d].

Os receptores geralmente tentam enviar um ACK, em resposta a alternativos segmentos de dados que recebem; esse envio se dá porque para muitas aplicações, se o receptor espera mais que o atraso esperado, ele pode enviar um ack de resposta como uma resposta normal para o emissor. Entretanto, quando o receptor recebe um segmento de dados fora de ordem, ele responde imediatamente com um ACK para retransmitir o segmento de dados perdido.

Quando um emissor recebe um ACK, ele realiza as seguintes determinações: pode determinar se o dado foi entregue ou não; pode determinar que o ACK é um Keepalive, utilizado para manter a linha ativa. No caso da não recepção do dado, o ACK determina que o receptor recebeu algum ou nenhum dado. No caso da recepção de algum dado, o ACK determina se novos créditos para envio dos dados serão permitidos. Quando um reconhecimento de ACK é recebido, sendo que não houve dados enviados e não há mais dados a serem enviados, o emissor interpreta o ACK como um ACK repetido. Esta condição indica que alguns dados foram recebidos fora de ordem, forçando com que o receptor envie o primeiro ACK, e que o segundo segmento de dados foi recebido fora de ordem, forçando assim com que o receptor envie o segundo ACK. Na maioria dos casos, o receptor receberá dois segmentos fora de ordem porque um dos segmentos foi descartado [CIS99a].

Quando um emissor de TCP detecta um segmento de dados descartado, ele retransmite o segmento. Então ele ajusta a taxa de transmissão que é a metade da existente antes do descarte detectado. Este é o comportamento conhecido como back-off ou slow down. Embora este comportamento seja apropriado para tratar congestionamento, problemas podem ocorrer quando múltiplas sessões concorrentes TCP encontram-se no mesmo roteador e todos os emissores TCP atrasam a transmissão dos pacotes ao mesmo tempo.

8.3.2.5 Como o roteador interage com TCP

No caso dos roteadores, eles podem manipular múltiplas sessões TCP concorrentes. Porque os fluxos de rede são adicionados aos poucos, existe uma grande probabilidade que quando o tráfego exceda o Transmit Queue Limit (TQL), ele irá exceder o limite. Entretanto, existe uma grande probabilidade que o tráfego excessivo seja temporário e que o tráfego não fique excessivo, exceto nos pontos nos quais ocorre o encontro entre tráfegos ou em roteadores das bordas [CIS99d].

Se um roteador descarta todo o tráfego que excede o TQL, como ocorre quando é usado o mecanismo tail drop, muitas sessões TCP irão simultaneamente se iniciar devagar. Conseqüentemente, o tráfego correrá temporariamente devagar ao extremo e então todos os fluxos irão se iniciar novamente devagar; esta atividade cria uma condição de sincronismo global.

Entretanto, se o roteador não descarta tráfego, como é o caso de mecanismos de filas como fair queueing (FQ) ou custom queueing (CQ), então o dado opera como se fosse armazenado na memória principal, degradando dramaticamente a performance do roteador.

No caso do RED, ele resolve o problema descrito acima levando uma sessão TCP ao retardo por vez, permitindo a utilização completa do bandwidth. Já no caso do mecanismo WRED, ele combina as características do algoritmo RED com o IP Precedence, de modo a prover uma manipulação de tráfego preferencial para pacotes com maior prioridade. O WRED pode seletivamente descartar tráfego de baixa prioridade quando a interface inicia o

congestionamento e provê diferentes características de performance para diferentes classes de serviços.

Para interfaces configuradas para utilizar Resource Reservation Protocol (RSVP), o WRED escolhe pacotes de outros fluxos a serem descartados em vez de fluxos RSVP. Também, o IP Precedence governa quais pacotes são descartados – o tráfego de baixa prioridade tem uma taxa de descarte maior que a dos de alta prioridade.

O WRED difere de outras técnicas para evitar congestionamento tal como estratégias de filas porque uma vez que ocorre congestionamento, no lugar de controlá-lo, ele procura se antecipar e evitá-lo.

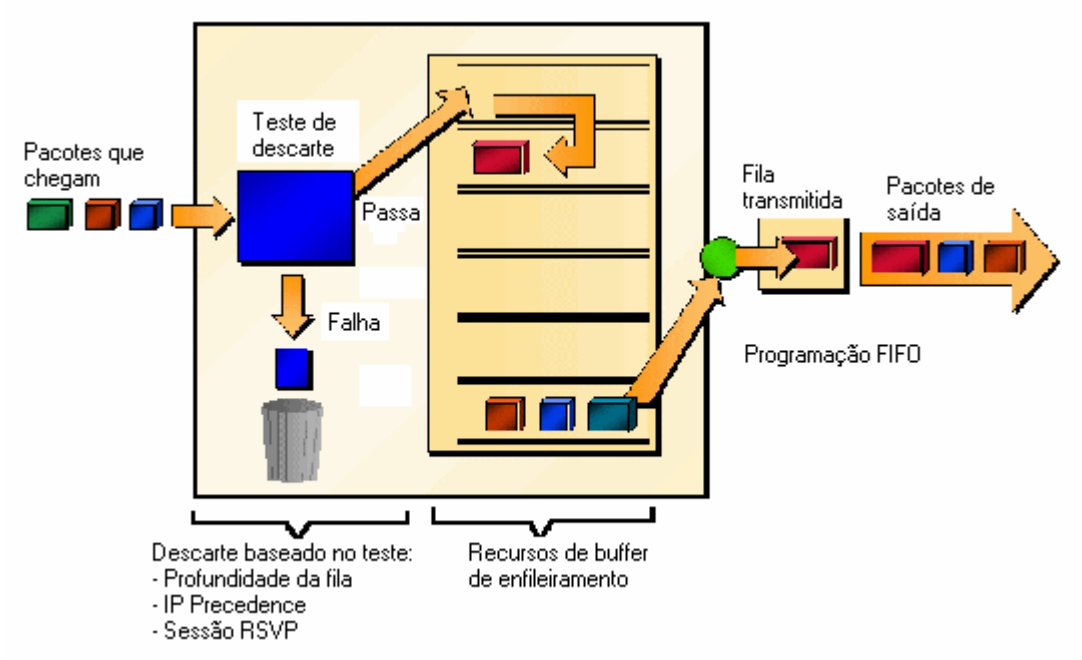

FIGURA 8.3.6 - Weighted Random Early Detection

8.3.2.6 Porque utilizar Weighted Random Early Detection?

O WRED faz a antecipação de detecção de congestionamento possível e provê para múltiplas classes de tráfego. Ele também protege contra a sincronização global. Por esta razão, o WRED é utilizado em qualquer interface de saída na qual é esperada uma ocorrência de congestionamento.

Entretanto, o WRED é geralmente usado em roteadores do core da rede, e não nos roteadores da borda. Roteadores de borda marcam precedência IP para pacotes quando eles entram na rede. O WRED usa esta precedência para determinar como tratar diferentes tipos de tráfego.

O WRED provê thresholds e pesos separados para diferentes IP Precendence, permitindo prover diferentes qualidades de serviços com descarte de pacotes para diferentes tipos de tráfego. Neste caso, o tráfego tradicional pode ser descartado mais freqüentemente que o tráfego do tipo premium durante períodos de congestionamento.

8.3.2.7 Como ele trabalha

Em períodos de congestionamento, quando os pacotes começam a ser descartados, o WRED avisa a fonte dos pacotes para decrescer sua taxa de transmissão. Se a fonte dos pacotes está utilizando TCP, ela decrescerá a taxa de transmissão de pacotes até que todos os pacotes alcancem o seu destino, que indica que o congestionamento não existe mais.

O WRED geralmente descarta pacotes seletivamente baseados no IP Pecedence. Os pacotes com o maior IP Precedence são descartados em menor proporção do que os de menor precedência. Assim, quanto mais alta a prioridade dos pacotes, mais alta a probabilidade de que os pacotes sejam entregues.

O WRED reduz as chances do tail drop através da seleção do descarte dos pacotes quando a interface de saída começa a mostrar sinais de congestionamento. Dropando alguns pacotes antecipadamente, em vez de esperar que a fila complete, o WRED evita um descarte numeroso de pacotes de uma só vez e minimiza as chances de sincronismo global. Assim, o WRED permite que a linha seja usada completamente.

Em soma, o WRED descarta estatisticamente mais pacotes de grandes usuários que pequenos. Assim, fontes de tráfego que geram mais tráfego são mais favoráveis a descarte que fontes com pequeno tráfego.

O WRED evita problemas de globalização que ocorrem quando o tail drop é utilizado como mecanismo para evitar congestionamento. A sincronização global se manifesta quando múltiplos hosts TCP reduzem suas taxas de transmissão em resposta ao descarte de pacotes, então – os hosts – aumentam suas taxas de transmissão novamente quando o congestionamento se reduz.

O WRED é útil quando o tráfego é do tipo TCP/IP. Com TCP, pacotes descartados indicam congestionamento, então a fonte dos pacotes reduz a taxa de transmissão. Com outros protocolos, a fonte dos pacotes talvez não venham a responder ou talvez retransmitam os pacotes descartados na mesma taxa. Assim, o descarte dos pacotes não diminui o congestionamento. O WRED trata o tráfego não IP com precedência 0, a precedência de menor prioridade. Neste caso, é mais provável que, em geral, o tráfego não IP seja descartado, ao invés do tráfego IP.

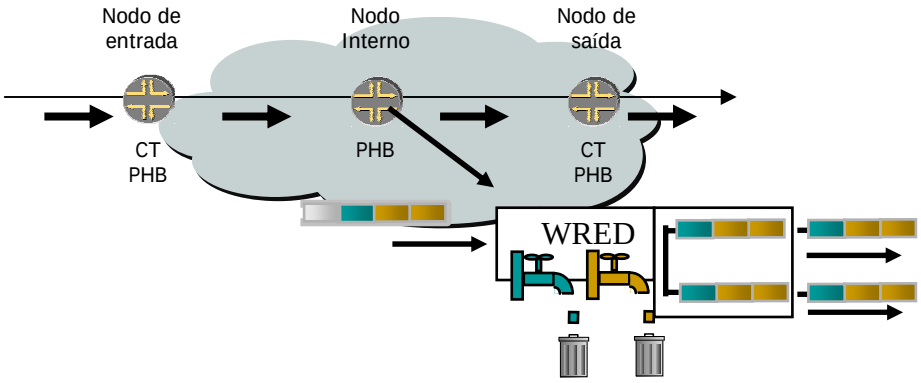

FIGURA 8.3.7 - Weighted RED

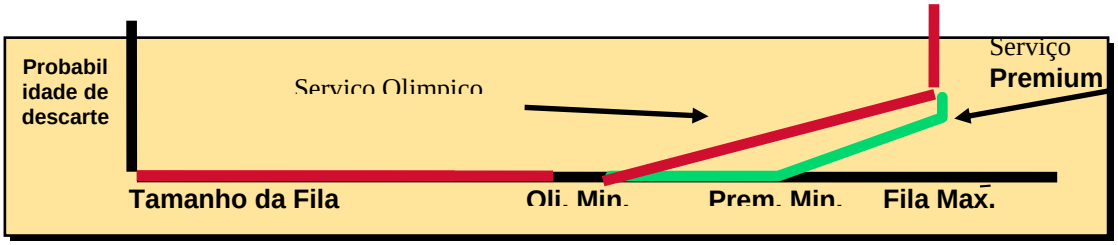

FIGURA 8.3.8 - Tráfego com Weighted RED

- O Weighted RED trabalha com múltiplos thresholds, um para cada classe de serviço. Serviços de baixa prioridade são descartados antes dos serviços de baixa prioridade, protegendo, assim, serviços do tipo premium.

8.3.2.8 O tamanho médio da fila

O roteador determina parâmetros, automaticamente, para utilizar nos cálculos do WRED. O tamanho médio da fila é baseado na média anterior e o tamanho corrente da fila. A formula é:

Média = (média_anterior * (1-2 –n)) + (tamanho_corrente_da_fila * 2 –n), onde n é o fator exponencial de peso configurado pelo usuário.

Para valores maiores que n, a média anterior é mais importante. Um fator maior leva facilmente ao pico e diminui o tamanho da fila. Não é provável que o tamanho médio da fila mude rapidamente, evitando mudanças drásticas de tamanho. O processo WRED inicia lentamente o descarte dos pacotes, mas continua descartando pacotes até que o tamanho atual da fila tenha sido menor que o minimum threshold. Esta mudança lenta da média possibilita acomodar rajadas temporárias de tráfego.

Nota: Se o valor de n chega ao máximo, o WRED não reage a congestionamentos. Os pacotes serão transmitidos ou descartados como se o WRED não tivesse efeito.

Para valores baixos de n, o tamanho médio da fila se assemelha muito ao tamanho da fila atual. A média resultante pode flutuar com as mudanças nos níveis de tráfego. Neste casos, o processo WRED responde rapidamente a filas longas. Uma vez que a fila alcança o minimum threshold, o processo para de descartar pacotes.

Se o valor de n for muito baixo, o WRED reage fortemente a rajadas de tráfego temporárias e descarta pacotes desnecessários.

8.4 Mecanismos de policiamento e conformação

Nesta seção descreveremos dois mecanismos para regular tráfego: o rate-limite do committed access rate (CAR) para policiamento de tráfego e o Generic Traffic Shaping (GTS) e Frame Relay Traffic Shaping (FRTS) para conformar o tráfego. Implementar estes mecanismos através da rede significa assegurar que um pacote, ou dado origem, siga o contrato estipulado. Tanto o policiamento como o mecanismo de conformação usam o descritor de tráfego para um pacote – indicado pela classificação do pacote – para assegurar o caminho e o serviço [CIS99c].

Policiais e conformadores geralmente identificam as violações no descritor de tráfego de maneiras idênticas. Eles geralmente diferem, entretanto, no modo como respondem à violação, por exemplo:

- Um policial tipicamente descarta o tráfego. (Por exemplo, o policial da taxa limite no CAR ou descarta o pacote ou reescreve seu IP Precedence, reconfigurando os bits do tipo de serviço do cabeçalho do pacote.)
- Um conformador tipicamente atrasa o tráfego excessivo, usando um buffer ou mecanismos de enfileiramento, para segurar pacotes e conformar o fluxo quando a taxa de dados da origem é mais alta do que o esperado. (Por exemplo, o GTS usa o Weighted fair queue para atrasar os pacotes para conformar o fluxo, e o FRTS usa o Priority Queue (PQ), Custom Queue (CQ), ou first-in, first-out (FIFO) para a mesma situação, dependendo de como é configurado.)

A conformação de tráfego e o policiamento podem trabalhar em conjunto. Por exemplo, um bom esquema de conformação de tráfego poderia ser utilizado para detectar fluxos com problemas de comportamento. Esta atividade é muitas vezes chamada de policiamento de fluxo de tráfego.

Esta seção descreve brevemente os mecanismos de policiamento e conformação de tráfego.

Uma vez que os mecanismos de policiamento CAR e conformação com FRTS e GTS trabalham com token bucket, este capítulo explica o seu funcionamento.

8.4.1 Leaky Bucket

O Leaky Bucket é um modelo usado para definir policiamento ou controle de admissão, em filas de ingresso. É importante relembrar que o leaky bucket não é uma estrutura de fila, mas um forma de descrever como a rede monitora e controla pacotes de usuários entrantes.

Para entendermos o modelo, a maneira mais fácil é pensar no modelo como um teste que é baseado em duas questões sobre o tráfego:

- O pacote deve ser permitido dentro da rede?
- Se sim, deveria ser dado a ele prioridade normal ou baixa?

O modelo leaky bucket é formado por três componentes:

- Token – um token chega no balde a uma velocidade definida pelo tráfego entrante. Cada pacote é representado no modelo leaky bucket como um único token. A taxa na qual os tokens preenchem o balde é variável. O dado, propriamente dito, do usuário não vai dentro do bucket.

- Bucket size – o tamanho do balde determina quantos tokens podem ser armazenados antes que o balde esteja cheio. Uma vez que o balde alcance sua capacidade, qualquer novo token representa pacotes que são noncompliant. Pacotes noncompliant, dependendo da situação, podem ser dropados ou marcados.

- Leak Rate – o leak rate (taxa de vazão) determina a taxa média aceitável de tokens para que esta se mantenha complacente. Se os tokens chegam a uma taxa aproximadamente igual, o balde nunca completará. Se os tokens chegam a uma taxa maior que a taxa de vazão, o balde se encontrará sempre no estado não complacente. Se os tokens chegam a uma taxa menor que a taxa de vazão, o balde se encontrará sempre vazio.

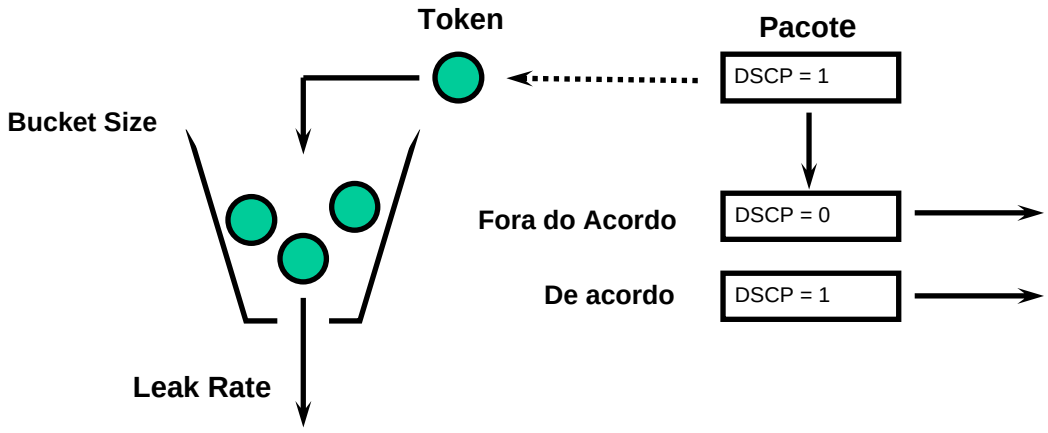

FIGURA 8.4.1 - Modelo Leaky Bucket

É difícil conceituar um balde cujo tamanho é configurado como um período de tempo. Outra forma de visualizar o mecanismo de policiamento é usar o teorema de chegada no tempo.

Cada pacote que chega pela rede é comparado ao tempo de chegada esperado. Se o pacote chegar antes, a penalidade é somada à penalidade total. A penalidade é baseada em

quão cedo o pacote chegou. Se o pacote chegar tarde, um crédito é dado e subtraído do total da penalidade. Se o total da penalidade exceder a configuração-limite o pacote é considerado não complacente e é descartado sem ser somado ao total da penalidade. O próximo pacote que chegar irá continuar a acumular penalidades e créditos.

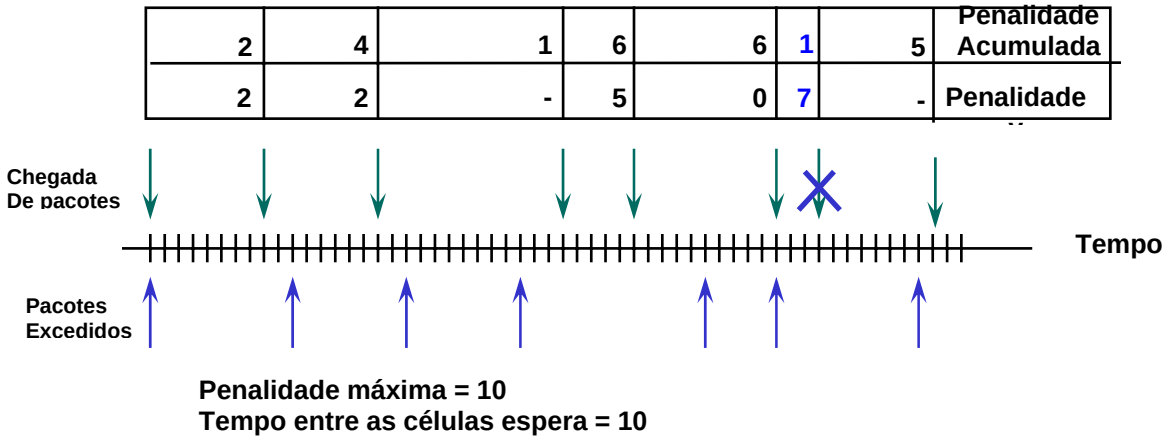

FIGURA 8.4.2 - Modelo baseado no tempo

8.4.2 Token Bucket

O token bucket é uma definição formal de um taxa de transferência. Ela possui três componentes: um tamanho da rajada, uma taxa resultante e um intervalo de tempo (Tc).

Aqui estão algumas definições dos termos:

- Taxa resultante – também chamada de committed information rate (CIR), ela especifica a quantidade de dados que podem ser enviados ou entregues por unidade de tempo em média.
- Tamanho da rajada – também chamada de Committed Burts (Bc) size, ela especifica em bits por rajada o quanto pode ser enviado dentro de uma dada unidade de tempo.
- Intervalo de tempo – também chamada de intervalo de medida, ela especifica o tempo em segundos por rajada.

Por definição, para cada múltiplo do intervalo integral, a taxa de bits da interface não pode exceder a taxa resultante. A taxa de bits, pode, entretanto, ser arbitrariamente mais rápida dentro do intervalo.

O token bucket é usado para gerenciar um dispositivo que regula o fluxo de dados. Por exemplo, o regulador poderia ser um policiador de tráfego, tal como CAR, ou um conformador de tráfego, tal como FRTS ou GTS. O próprio token bucket não tem políticas de descarte ou priorização. Ao contrário, um token bucket descarta tokens e deixa para o fluxo o problema de gerenciar sua fila de transmissão, se o fluxo sobrecarrega o regulador.

De acordo com a metáfora token bucket, os tokens são colocados dentro do bucket numa certa taxa. O próprio bucket tem uma capacidade especificada. Se o bucket preenche sua capacidade, os novos tokens que chegam são descartados. Cada token tem a permissão de ser enviado da origem em um certo número de bits na rede. Para transmitir um pacote, o regulador deve remover do bucket um número de tokens igual a representação do tamanho do pacote [CIS99c].

Se não existem tokens o suficiente no bucket para enviar um pacote, o pacote deve esperar até que o bucket os tenha suficientemente ou o pacote será descartado. Se o bucket já está cheio de tokens, os tokens que entram sobrecarregam e não são disponíveis para futuros pacotes. Assim, uma rajada mais larga de uma origem na rede é proporcional ao tamanho do bucket.

Note-se que o mecanismo de token bucket usado para conformação de tráfego tem um token bucket e um buffer de dados, ou fila; se não tiver um buffer de dados, ele poderia ser um policiador. Para conformação de tráfego, pacotes que chegam e que não podem ser enviados imediatamente são atrasados no buffer de dados.

Para a conformação de tráfego, um token bucket permite rajadas de modo limitado. Ele garante que a rajada é limitada, uma vez que o fluxo nunca envia mais rápido que a capacidade do token bucket, mais o intervalo de tempo, dividido pela taxa estabelecida a qual os tokens são colocados no bucket. Ele também garante que as taxas de transmissão de longo tempo não irão exceder a taxa estabelecida que os tokens são colocados no bucket.

8.4.3 Policiamento com Committed Access Rate

O CAR incorpora uma característica de taxa-limite para policiamento de tráfego, somado à sua característica de classificação. Esta característica gerencia o policiamento do acesso ao bandwidth da rede, assegurando que o tráfego, de acordo com os parâmetros de taxa especificados seja transmitido, enquanto descarta pacotes que excedam a quantidade de tráfego ou transmitindo-os com uma prioridade diferente [CIS99c].

As funções do limitador de taxa são as seguintes:

- Permite controlar a taxa máxima de tráfego transmitido ou recebido em uma interface;
- Dá a habilidade de definir agregados de nível 3 ou taxa-limite de bandwidth de ingresso ou saída e para especificar políticas de manipulação de tráfego quando o tráfego está conforme ou excede taxas- limites especificadas.

Nota:

- A taxa limitada de bandwidth agregado está associada a todos os pacotes em uma interface ou sub-interface;

- A taxa limite de bandwidth granular está associada a um tipo particular de tráfego baseado em precedência, endereço MAC ou outros parâmetros;
- O CAR é configurado, na maioria das vezes, em interface na borda da rede para limitar o tráfego na ou fora da rede.

8.4.4 Como funciona

O CAR examina o tráfego recebido na interface ou um subset deste tráfego selecionado por critérios da lista de acesso. Ele então compara a taxa de tráfego a um token bucket configurado, e age de acordo com os resultados. Por exemplo, o CAR descartará o pacote ou reescreverá o IP Precedence, reconfigurando os bits type-of-service (ToS).

Esta seção explica os seguintes aspectos do limitador de taxa CAR:

- De acordo com critérios
- Limite de taxa
- Ações conformes ou excedentes
- Múltiplos policiamentos de taxa

8.4.5 Critérios

Políticas de taxa podem ser associadas com um dos itens seguintes:

- Interface de entrada
- Todo o tráfego IP
- IP Precedence (definido por uma lista de acesso de limite de taxa)
- MAC address (definido por uma lista de acesso de limite de taxa)
- Lista de acesso IP (padrão e estendida)

O CAR provê ações de configuração, tal como transmitir, descartar ou configurar precedência quando o tráfego é conforme e excede a taxa-limite.

Nota Associações por lista de acesso IP ocupam mais tempo de processamento que associar baseado em outros critérios.

8.4.6 Limites de taxa

O CAR propaga rajadas. Ele não realiza smoothing ou conformação de tráfego, e por isso ele não bufferiza e não adiciona atraso. O CAR é altamente otimizado para rodar em links de alta performance.

Os limites de taxa do CAR podem ser implementados em interfaces ou sub-interfaces de entrada e saída, incluindo Frame Relay e ATM.

Limites de taxa definem qual pacote é conforme ou excede a taxa definida baseada nos seguintes parâmetros:

- Taxa média – A taxa média determina a taxa de transmissão média em um tempo longo. O tráfego que cai nesta taxa está sempre conforme;
- Tamanho normal da rajada – O tamanho normal da rajada determina quão longa a rajada de tráfego pode ocorrer, antes de algum tráfego exceder a taxa-limite.
- Tamanho da rajada excedente – O tamanho da rajada excedente determina quão longo o tráfego da rajada pode ocorrer antes que todo o tráfego exceda a taxa-limite. O CAR provê gerenciamento de descarte entre a rajada excedida e a extensão dos parâmetros da rajada excedida. O tráfego que cai, entre o tamanho normal de rajada e o tamanho de rajada excedente, excede o limite da taxa com a probabilidade de aumentar com o aumento do tamanho da rajada.

O número máximo de tokens que um bucket pode conter é determinado por um tamanho normal de rajada, configurado para o token bucket.

Quando o limite de taxa do CAR é aplicado a um pacote, o CAR remove do bucket tokens equivalentes ao número de bytes do pacote. Se um pacote chega e o número de tokens existentes no token bucket padrão é menor que o tamanho de bytes do pacote, o extended burst é aplicado.

Quando o extended burst é configurado permite-se a utilização de tokens emprestados para poder transmitir o pacote. Esta capability existe para evitar comportamentos do tipo tail-drop, e, ao invés, engajar comportamentos como o do Random Early Detection (RED).

O extended burst trabalha da seguinte forma: se um pacote chega e necessita pedir emprestado n tokens porque o token bucket contêm menos tokens que os necessários para um pacote, então o CAR compara os dois seguintes valores:

- Valores dos parâmetros extended burst;
- Debito composto. O Débito composto é computado através da soma sobre *a* e *i*

— *i* indica o número de tokens a serem emprestados desde a última vez que o pacote foi descartado.
— *a* indica o atual valor de débito do fluxo depois que o pacote é enviado.

Se um débito é maior que o valor de rajada estendida, a ação de exceção do CAR é efetivada. Depois que o pacote é descartado, o débito composto é efetivado para 0. O CAR irá computar um novo valor de débito para o atual débito para o próximo pacote que necessitar pedir tokens emprestados.

Se o atual débito é maior que o limite estendido, todos os pacotes serão descartados até que o débito atual seja reduzido, através do acúmulo de tokens no token bucket.

Descartar pacotes não conta contra qualquer taxa ou limite de rajada. Quando um pacote é descartado, nenhum token é removido do token bucket.

Testar o tráfego do Transmission Control Protocol (TCP) utilizando valores para rajadas normais e estendidas é alguns segundos pior do que o tráfego a uma taxa média. Se uma taxa média é de 10 Mbps, então um tamanho normal de rajada de 10 à 20 Mbps e um tamanho de rajada excessiva de 20 à 40 Mbps seria apropriado.

8.4.7 Ações conformes ou excedentes

O CAR utiliza um token bucket, assim o CAR pode passar rajadas temporárias que excedam a taxa-limite [CIS99c].

Uma vez que um pacote tenha sido classificado como conforme ou excedente a uma taxa-limite, o roteador realiza uma das seguintes ações no pacote:

- Transmit – o pacote é transmitido;
- Drop – o pacote é descartado;
- Set precedence and transmit – os bits IP Precedence (ToS) no cabeçalho do pacote são reescritos, e o pacote é então transmitido;
- Continue – o pacote é avaliado usando uma nova taxa de policiamento de taxa-limite. Se não existe outra taxa de policiamento, o pacote é transmitido.

8.4.8 Múltiplos policiamentos de taxa

Um único policiamento de taxa do CAR inclui informações sobre a taxa- limite, ações de conformidade e ações que excedem. Cada interface pode ter múltiplos policiamentos de taxa de CAR correspondendo a diferentes tipos de tráfego. Quando existem múltiplas taxas de policiamento, o roteador examina cada policiamento, de modo a encontrar uma associação com o pacote. Se nenhuma associação é localizada, a ação default é transmitir.

Políticas de taxa podem ser independentes: cada política de taxa fecha com um tipo de tráfego. Alternativamente, políticas de taxa podem ser cascateadas: um pacote pode ser comparado a múltiplas políticas de taxa em sucessão.

8.4.9 Conformação de tráfego

Vamos descrever nesta seção dois tipos de conformadores de tráfego: o GTS e o FRTS. Os métodos são similares na sua implementação, usam diferentes tipos de filas para armazenar e conformar o tráfego. Em particular, o código que determina quando existe crédito suficiente token bucket para um pacote ser enviado, ou quando o pacote deve ser atrasado, é comum para ambos. Se um pacote esta atrasado, o GTS usa Weighted Fair Queue para assegurar o tráfego atrasado. Já o FRTS usa o Custom Queue ou o Priority Queue, dependendo da configuração.

Esta seção descreve como o conformador de tráfego trabalha. Ela inclui as seguintes subseções:

- Conformador de tráfego
- Generic Traffic Shaping (GTS)
- Frame Relay Traffic Shaping (FRTS)

8.4.9.1 O conformador de tráfego

O conformador de tráfego permite o controle de tráfego que sai pela interface, de modo a associar a velocidade do fluxo com a interface remota e garantir que o tráfego seja conforme ao policiamento contratado. Assim, o tráfego de um perfil particular pode ser conformado para encontrar os requisitos do downstream, eliminando possíveis gargalos na topologia [CIS99c].

8.4.9.2 Por que utilizar conformação de tráfego?

A primeira razão para se utilizar conformação de tráfego é o controle de acesso para com o bandwidth disponível, assegurando que o tráfego seja conforme ao policiamento estabelecido para ele, e para regular o fluxo do tráfego, de modo a evitar congestionamentos que possam ocorrer quando o tráfego excede a velocidade da interface remota. Por exemplo:

- Controle de acesso ao bandwidth quando o policiamento dita que a taxa de uma dada interface não deveria, na média, exceder uma certa taxa, embora a taxa de acesso exceda a velocidade.
- Configurar conformação de tráfego na interface tendo uma rede com diferentes taxas de acesso. Suponha que os links Frame Relay tenha 128 kbps e 256 kbps. Enviar pacotes a 256 kbps pode causar falha na aplicação.

O conformador de tráfego previne perda de pacotes. O seu uso é especialmente importante em redes Frame relay porque o switch não pode determinar quais pacotes têm precedência, além de quais pacotes devem ser descartados quando o congestionamento ocorre. Em soma, é de importância crítica para o tráfego do tipo real-time, tal como Voz sobre Frame relay cuja latência é limitada, limitando assim a quantidade de tráfego e perda de tráfego em links de redes de dados em qualquer tempo, mantendo o dado em roteadores que fazem a garantia. A retenção do dado em roteadores permite que o roteador priorize o tráfego de acordo com a garantia. (A perda de pacotes pode ser resultante da conseqüência de aplicações real-time e interativas.)

8.4.9.3 O conformador de tráfego e a taxa de transferência

O conformador de tráfego limita a taxa de transmissão de dados através de:

- Uma taxa específica configurada;
- Uma taxa derivada, baseada no nível de congestionamento.

Como mencionado, a taxa de transferência depende de três componentes que constituem o token bucket: tamanho da rajada, taxa resultante, intervalo de tempo. A taxa resultante é igual ao tamanho da rajada dividido pelo intervalo.

Quando o conformador de tráfego é habilitado, a taxa de bits da interface não irá exceder a taxa resultante sobre qualquer intervalo múltiplo do integral. Em outras palavras, durante todo o intervalo, um tamanho máximo de rajada pode ser transmitido. Dentro do intervalo, entretanto, a taxa de bit pode ser mais rápida que a taxa resultante em qualquer tempo dado.

Uma variável adicional aplicada ao conformador de tráfego: Be size. O tamanho da rajada excedida corresponde ao número de bits não enviados – fora do CIR – que ainda são aceitos pelo Frame Relay, mas marcados para descarte, se necessário.

Em outras palavras, o Be size permite mais que o tamanho de rajada possa ser enviada durante um intervalo de tempo em certas situações. O switch permitirá que pacotes além da rajada excessiva passem, mas serão marcados com o bit Discard Eligible (DE). Quando o Be size é igual a zero, a interface não envia mais que o tamanho da rajada em todo o intervalo, alcançando uma taxa média não maior que a taxa resultante. Entretanto, quando o Be size é maior que 0, a interface pode enviar Bc+Be bits em uma rajada, se anteriormente a quantidade máxima não foi transmitida. Quando, menos que o tamanho da rajada é transmitido durante um intervalo, o número de bits restantes, até o tamanho da rajada excedida, pode ser utilizado para transmitir mais que o tamanho da rajada em um próximo intervalo.

8.4.9.4 O bit Discard Eligible

Podemos especificar qual pacote Frame Relay tem baixa prioridade ou baixa sensibilidade ao tempo e será o primeiro a ser descartado quando um Frame Relay estiver congestionado. O mecanismo que permite um Frame Relay identificar tal pacote é o bit DE.

Podemos definir listas de DE que identificam características de pacotes a serem descartados, e podemos também especificar grupos de DE para identificar um DLCI que é afetado.

Podemos especificar uma lista DE, baseada em protocolos ou interfaces, e em características, tal como fragmentação de pacotes, um TCP especifico ou porta User Datagram Protocol (UDP), um número de lista de acesso, ou tamanho de pacote.

8.4.9.5 Diferenças entre Generic Traffic Shaping e Frame Relay Traffic Shaping

Como mencionado, tanto o GTS como o FRTS são similares nas suas implementações, compartilhando o mesmo código e estrutura de dados, mas eles diferem nos tipos de filas que são utilizados.

Existem duas diferenças entre o GTS e o FRTS:

- O FRTS suporta compartilhamento baseado em DLCI, enquanto o GTS é configurado por interface ou sub-interface.
- Para GTS, a fila utilizada é uma Weighted Fair Queue (WFQ). Para FRTS, a WFQ não é suportada; ao invés, a fila pode ser CQ, PQ ou FIFO.

TABELA 8.4.1 - Comparação entre FRTS e GTS

	FRTS	**GTS**
Linha de comandos	• Parâmetros de classes • Parâmetros aplicados a todos os canais virtuais ou a interface • Sem comandos de grupo	• Comandos aplicados por sub-interfaces • Comandos de grupo
Filas suportadas	• CQ, PA, FCFS por VC	• WFQ por sub-interface

8.4.9.6 Conformador de tráfego e enfileiramento

O conformador de tráfego conforma o tráfego acima da taxa configurada em uma fila.

Quando um pacote chega em uma interface de transmissão, acontece o seguinte:

- Se uma fila está vazia, o pacote que chega é processado pelo conformador de tráfego.
- Se possível, o conformador de tráfego envia o pacote;
- Se não, o pacote é colocado em uma fila.
- Se a fila não está vazia, o pacote é aí colocado.

Quando existem pacotes em uma fila, o conformador de tráfego remove o número de pacotes que ele puderem um dado intervalo de tempo.

8.4.9.7 O Generic Traffic Shaping

O GTS conforma o tráfego reduzindo o fluxo do tráfego de saída para evitar congestionamento, levando o tráfego a uma taxa de bit particular usando mecanismo de token bucket.

O GTS aplica-se a uma interface e a lista de acesso pode ser usada para selecionar o tráfego a ser conformado. Ele trabalha com uma variedade de tecnologias de nível 2, incluindo Frame Relay, ATM, Switched Multimegabit Data Service (SMDS), e Ethernet.

Em sub-interfaces Frame Relay, o GTS pode ser configurado para adaptar dinamicamente a uma bandwidth disponível pela integração de sinais BECN, ou configurado simplesmente para conformar a uma taxa. O GTS pode também ser configurado em uma interface ATM para responder à sinalização RSVP sobre uma configuração ATM de PVCs.

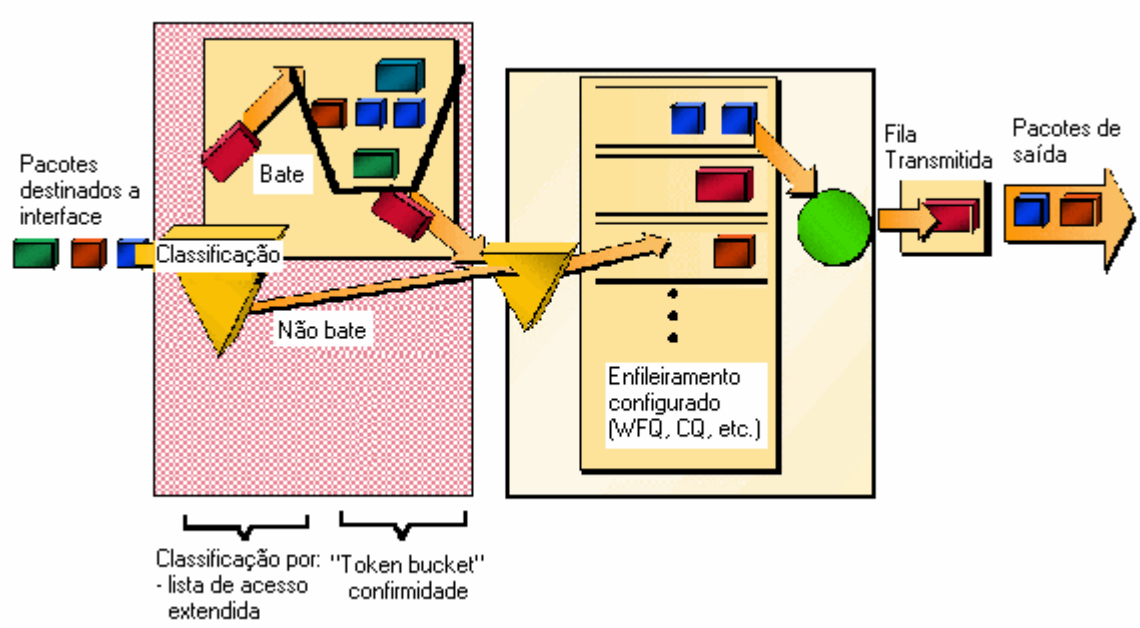

FIGURA 8.4.3 - Generic traffic shaping

8.4.10 O Frame Relay Traffic Shaping

O FRTS foi construído sobre tecnologias como Forward Explicit Congestion Notification (FECN), BECN e DE, de modo a prover escalabilidade e performance a redes Frame Relay, aumentando a densidade de circuitos virtuais e aumentando o tempo de resposta.

Tanto o GTS como o FRTS podem eliminar gargalos na rede Frame Relay com conexões de alta velocidade no site central e conexões de baixa velocidade nos escritórios. Podemos configurar uma taxa de pico para limitar o tráfego de saída, limitando a taxa de dados no VC do site central.

Podemos também definir PW, CQ e WFQ no VC ou sub-interface. Usando estes métodos de enfileiramento torna-se possível uma granulosidade na priorização e enfileiramento de tráfego, provendo mais controle sobre o fluxo de tráfego em um VC individual. Se combinarmos CQ com enfileiramento por VC e limitar a taxa, podemos habilitar VCs Frame Relay a carregar múltiplos tipos de tráfego tais como IP, SNA e IPX com bandwidth garantido para cada tipo de tráfego.

Usando a informação contida no pacote marcado com BECN recebido da rede, o FRTS pode também, dinamicamente, parar suavemente o tráfego. Com paradas baseadas em BECN, pacotes são segurados nos buffers dos roteadores para reduzir o fluxo de dados do roteador na rede Frame Relay. A parada é realizada por VC e a taxa de transmissão é ajustada baseando-se no número de pacotes marcados com BECN recebidos.

Assim, poderíamos integrar FRST com ATM ForeSight para controle de congestionamento fim a fim em redes multiserviço.

8.4.10.1 Taxas derivadas

Em redes Frame relay, os bits BECNs e FECNs dentro do frame Frame relay indicam congestionamento.

O FECN é gerado quando o dado é enviado de uma interface congestionada; indica para o DTE que um congestionamento foi encontrado. O tráfego é marcado com BECN se a fila para a direção oposta está próxima ao threshold FECN.

O BECN notifica o emissor para decrescer a taxa do transmissor. Se o tráfego é de um lado apenas (um tráfego multicast), não existe tráfego reverso com o BECN para notificar o emissor para ir devagar. Assim, quando um DTE recebe um FECN, ele primeiro determina se ele está enviando algum dado de retorno. Se ele retorna dados, estes dados serão marcados com um BECN para o outro DTE. Entretanto, se o DTE não enviar dados, o DTE pode enviar uma mensagem Q,922 TEST RESPONSE com o bit configurado para BECN.

Quando uma interface configurada com conformação de tráfego recebe um BECN, ele imediatamente decresce sua taxa máxima em uma grande quantia. Se, depois de alguns intervalos, a interface não receber outro BECN e o tráfego estiver esperando na fila, a taxa máxima aumenta gradativamente. O ajuste dinâmico da taxa máxima é chamada de taxa derivada.

8.4.10.2 Restrições

Pelo método o qual o FRTS é implementado, ele deve ser utilizado em tráfego real-time. Se o FRTS for utilizado para tráfego em rajadas para Be rate, o roteador deve esperar um período de tempo antes de transmitir novamente. Em certas condições, este tempo pode ser de até 900 milisegundos, uma quantia de tempo inaceitável para tráfego em tempo real.

8.5 Mecanismos de eficiência de link

Encontramos dois tipos de mecanismos de eficiência de link – o Link Fragmentation and Interleaving (LFI) para Multilink Point-to-Point Protocol (MLP), e o Compressed Real-Time Protocol (CRTP) header — o qual trabalha com enfileiramento e conformação de tráfego para aumentar a eficiência e prevenção dos níveis de serviço de aplicações.

Esta seção dá uma breve introdução dos seguintes mecanismos de eficiência:

- Link Fragmentation and Interleaving
- Compressed Real-Time Protocol Header

8.5.1 O Link Fragmentation and Interleaving

O tráfego interativo, tal como Telnet e Voz sobre IP, é suscetível ao aumento de latência quando a rede processa um grande número de pacotes, tal como File Transfer Protocol (FTP) da LAN para a WAN. O atraso nos pacotes é especialmente significante quando os pacotes FTP são enfileirados em links lentos na WAN. Para resolver problemas de atraso em links de banda pequena, torna-se necessário um método para fragmentar pacotes grandes e então enfileirá-los.

O LFI reduz o atraso em links de pequena velocidade quebrando grandes datagramas e incluindo os pacotes que necessitam de baixo atraso entre os pequenos pacotes resultantes da fragmentação dos pacotes. No caso do LFI, ele se utiliza da implementação do MLP, o qual suporta fragmentação e especificações de sequenciamento de pacotes no RFC 1717.

O LFI permite reserva de filas de modo que fluxos de Real-Time Protocol (RTP) possam ser mapeadas dentro de uma fila de mais alta prioridade.

8.5.1.1 Como funciona

Para entender como o LFI usando MLP funciona, é melhor entender o problema sobre o destino do endereçamento. O atraso fim a fim até o destino para pacotes de tempo real, especialmente voz, é de 150 a 200 milisegundos (ms). As técnicas de transmissão de datagramas baseados em IP para transmissão de áudio não endereçam de maneira adequada o problema imposto pelo limite de banda e o atraso de 150 ms.

O inaceitável atraso de fila para pacotes pequenos de tempo real existe mesmo com o uso de características de QoS como: Resource Reservation Protocol (RSVP) e Weighted Fair Queueing (WFQ), e o uso de algoritmos de compressão de voz, tal como Compressed Encoding for Linear Prediction (CELP), o qual reduz a taxa de 64 Kbps para 8 Kbps.

Mesmo com estes problemas,o atraso de tempo real continua a existir porque o overhead por pacote é muito grande e os large maximum transmission units (MTUs) são necessários para produzir uma vazão aceitável para uma transmissão eficiente.

Um MTU de 1500 bytes leva 215 ms para atravessar uma linha de 56 kbps, o que excede o atraso até o destino. Deste modo, para limitar o atraso de pacotes de tempo real em links de baixa velocidade – como 56 kbps e 64 Kbps – um método para fragmentação de pacotes grandes e enfileiramento de pacotes menores entre fragmentos de pacotes maiores é necessário. O MLP ajuda a resolver este problema através do LFI.

O MLP provê um método de dividir, recombinar e seqüenciar datagramas através de múltiplos links lógicos. O esquema LFI é relativamente simples: datagramas grandes são encapsulados e fragmentados para pacotes de tamanhos pequenos o suficiente para satisfazer o atraso requisitado do tráfego sensível ao atraso; os pacotes pequenos sensíveis ao atraso não são encapsulados, mas são inseridos entre fragmentos de um datagrama grande.

O MLP permite que os pacotes fragmentados sejam enviados ao mesmo tempo sobre múltiplos links ponto a ponto para o mesmo endereço remoto. O MLP provê largura de banda em demanda e reduz a latência da transmissão através de links WAN.

A figura abaixo mostra o mix de tráfego destinado para a interface. Baseado na sua classificação, os pacotes que chegam são agrupados dentro de filas. Depois que os pacotes estiverem na fila, o jumbogram é fragmentado dentro de pequenos pacotes em preparação à inclusão dos pacotes de voz, sensíveis ao tempo.

Porque o WFQ é configurado por interface, os pacotes de cada fila – os fragmentos do pacote jumbogram e os pacotes de voz IP – são intercalados e programados (baseados no seu peso) para serem transmitidos na fila da interface de saída.

De modo a assegurar a ordem correta de transmissão e remontagem, o LFI adiciona múltiplos cabeçalhos aos datagramas fragmentados depois que os pacotes são desenfileirados e prontos para transmitir.

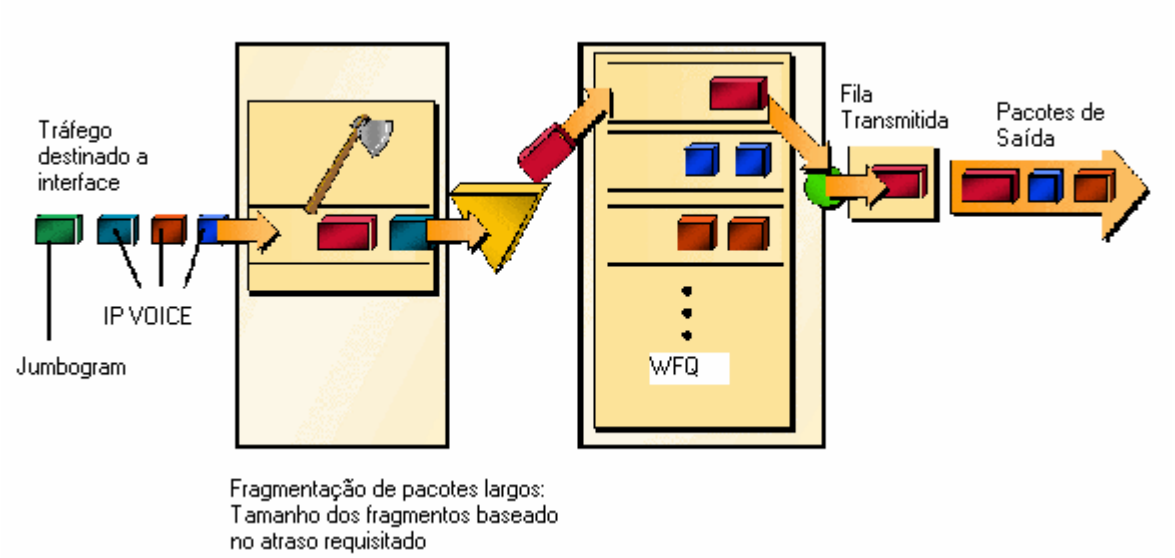

FIGURA 8.5.1 - Link Fragmentation and Interleaving

8.5.2 O Compressed Real-Time Protocol Header

O RTP é um padrão de protocolo Internet para transporte de dados em tempo real. Ele provê funcionalidades de transporte de rede fim a fim para aplicações que suportam áudio, vídeo, ou simulam dados sobre serviços de rede multicast ou unicast.

O RTP provê suporte a conferências em tempo real de qualquer tamanho dentro da Internet. Este suporte inclui identificação da fonte e suporte a gateways, tal como bridges de áudio e vídeo, bem como tradutores de multicast para unicast. O RTP oferece feedback de QoS de receptores para grupos de multicast, bem como suporte para a sincronização de diferentes fluxos.

O RTP inclui uma porção de dados e um porção de cabeçalho. A porção dados do RTP é um protocolo pequeno que provê suporte para propriedades de tempo real de aplicações, tal como mídia contínua, incluindo reconstrução de timing, detecção de perda e identificação de conteúdo.

A porção do cabeçalho do RTP é consideravelmente grande. Como mostra a seguinte figura, o tamanho mínimo do cabeçalho RTP de 12 bytes, combinado com 20 bytes do cabeçalho IP (IPH) e 8 bytes do cabeçalho UDP, criam um cabeçalho de 40 bytes IP/UDP/RTP. Para cada carga útil comprimida de aplicações de áudio, o RTP tipicamente tem 20 bytes para 160 bytes de carga útil. Dado o tamanho da combinação do cabeçalho IP/UDP/RTP, torna-se ineficiente transmitir o cabeçalho IP/UDP/RTP sem comprimi-lo.

De modo a evitar o consumo desnecessário de banda disponível, a característica de compressão do cabeçalho RTP – referenciada como CRTP – é usado por default.

8.5.2.1 Como funciona

O CRTP comprime o cabeçalho IP/UDP/RTP em um pacote de 40 bytes para aproximadamente 2 a 5 bytes.

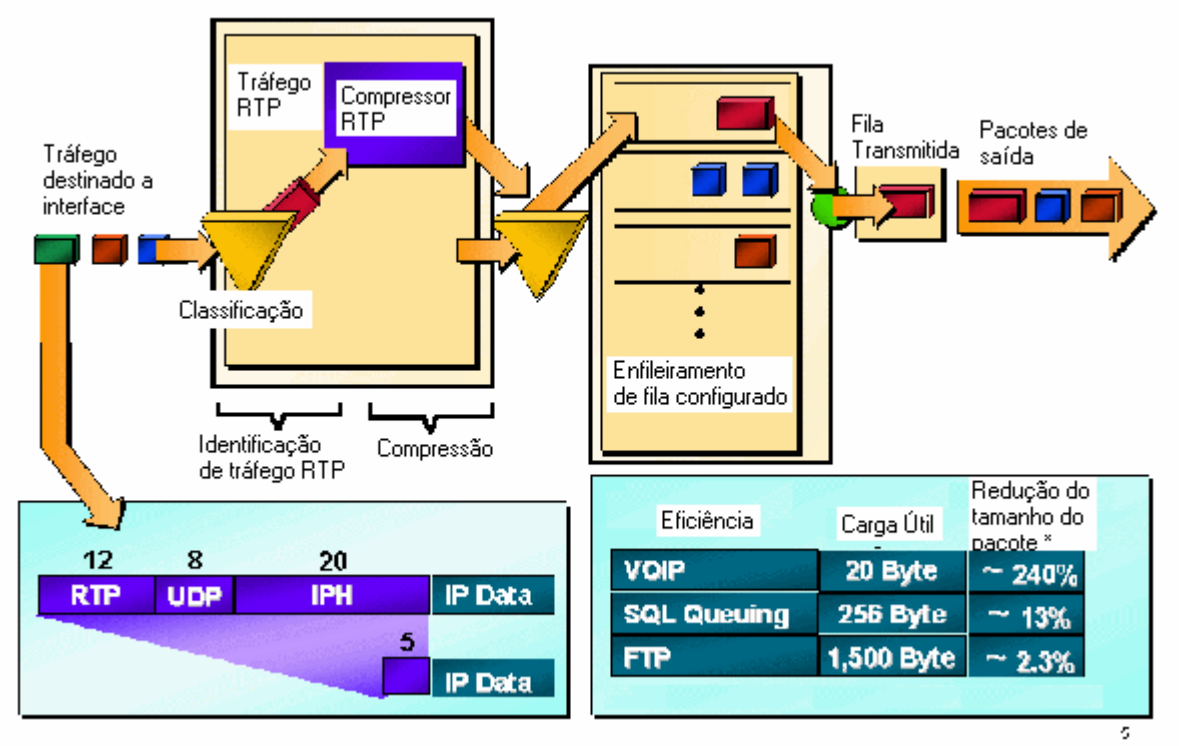

FIGURA 8.5.2 - RTP Header Compression

8.5.2.2 O Compressed Real-Time Protocol Header

O CRTP é um esquema de compressão similar ao TCP header compression (RFC 1144).

8.5.2.3 Por que usar o Compressed Real-Time Protocol Header?

O CRTP reduz o overhead na linha para tráfego multimídia RTP resultando em um igual redução no atraso; o CRTP é especialmente benéfico quando a carga útil do RTP é pequena, por exemplo, para carga útil de áudio comprimido de 20 a 50 bytes.

Podemos usar o CRTP para media em demanda e serviços interativos, tal como telefone via Internet. Como o RTP, o CRTP provê suporte a conferências em tempo real de qualquer tamanho na Internet.

8.6 Mecanismos de sinalização

De um modo geral, a sinalização QoS é uma forma de comunicação de rede que provê um caminho para uma estação final ou nodo de rede se comunicar com, ou sinalizar, seus vizinhos para requisitar manipulação especial de certo tráfego. Os requisitos de QoS fim a fim que todos os elementos no caminho da rede devem entregar, deve ser coordenado com sinalização QoS [CIS99b].

Muitas soluções de sinalização QoS provêem QoS em alguns lugares da infra-estrutura; entretanto, acaba existindo um escopo limitado através da rede. Para alcançar QoS fim a fim, a sinalização deve estar em toda a rede.

Uma rede pode alcançar QoS fim a fim, por exemplo, usando parte do cabeçalho do pacote IP para requisitar manipulação de priorização ou tráfego sensível ao tempo.Pelo fato do IP estar em todo o lugar, o QoS tira vantagem de prover sinalização fim a fim, utilizando sinalização in-band (IP Precedence, 802.1p) ou out-of-band (RSVP) para indicar que um serviço com QoS em particular é desejado para uma classificação de tráfego particular – IP Precedence para sinalização de QoS diferenciado e RSVP para QoS garantidos.

8.6.1 IP Precedence

Como mostra a figura abaixo, o IP Precedence utiliza os três bits precedentes do campo ToS no cabeçalho do IPv4 para especificar a classe de serviço para cada pacote. O tráfego pode ser particionado em até seis classes usando o IP Precedence. As tecnologias de enfileiramento através da rede podem então utilizar esta sinalização para prover a manipulação apropriada.

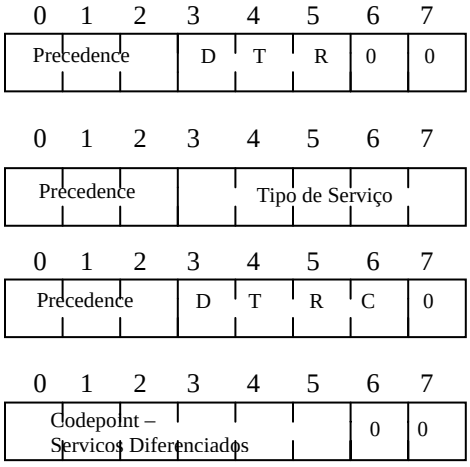

FIGURA 8.6.1 - Histórico do campo IP

Podem ser utilizadas características tais como policy-based routing (PBR) e committed access rate (CAR) para configurar precedência baseada em classificações de lista de acesso. O uso desta característica permite uma flexibilidade considerável, incluindo precedência por aplicação ou usuário, ou por subrede destino e origem. Tipicamente, a implementação desta tecnologia ocorre na borda da rede. O IP Precedence pode também ser configurado no host ou na rede do cliente; entretanto, o IP Precedence pode ser reescrito pelo policiamento dentro da rede [CIS99b].

O IP Precedence habilita a existência de classes de serviço usando mecanismos de enfileiramento existentes na rede, tais como WFQ e WRED, sem nenhuma mudança para as aplicações já existentes e sem requisitos complicados de rede.

8.6.2 Resource Reservation Protocol

O RSVP é o primeiro protocolo padronizado na indústria para configuração dinâmica de QoS fim a fim através de redes heterogêneas. O RSVP, o qual roda sobre IP, permite que uma aplicação dinamicamente reserve bandwidth na rede. Usando RSVP, aplicações podem requisitar um certo nível de QoS para um fluxo de dados através da rede.

O RSVP não realiza seu roteamento próprio; ao invés disso, ele se utiliza de outros protocolos para determinar onde deve passar as requisições de reserva de banda. Como o roteamento muda seus caminhos para adaptar as mudanças da topologia, o RSVP adapta-se para o novo caminho. Esta modularidade não inibe o RSVP de usar outros serviços de roteamento. O RSVP provê operação transparente através de roteadores que não suportam RSVP [CIS99b].

O RSVP trabalha em conjunto com mecanismos de enfileiramento. Quando o RSVP requisita um QoS em particular, mecanismos como Weights Fair Queueing (WFQ) ou Weighted Random Early Detection (WRED) são responsáveis em implementar a reserva. Dois tipos de reservas dinâmicas podem ser feitas: controle de carga e serviços de taxa garantida.

A primeira característica do RSVP é a escalabilidade. O RSVP torna-se escalar utilizando a escalabilidade herdada pelo multicast, o qual utiliza requisições de reserva orientadas pelo receptor. Embora o RSVP seja projetado especialmente para aplicações multicast, ele pode também realizar reservas unicast, embora ele não seja tão escalar com um grande número de reservas unicast.

O RSVP é uma característica importante de QoS, mas ele não resolve todos os problemas relacionados ao QoS, e eles impõem detalhes como tempo necessário para configuração de reserva fim a fim.

8.6.3 Como funciona

Os hosts e roteadores usam o RSVP para enviar requisitos de QoS para roteadores, ao longo do caminho da stream de dados, e para manter o estado de roteadores e host para prover o serviço requisitado, geralmente bandwidth e latência. O RSVP usa uma conhecida taxa de dados, a maior quantidade de dados que o roteador pode manter na fila, e o QoS mínimo para determinar a reserva do bandwidth [CIS99b].

Um host usa RSVP para requisitar um serviço com QoS específico da rede, com o interesse no stream de dados para aplicação. O RSVP requisita um QoS particular, mas a reserva acontece mesmo, pelos mecanismos de enfileiramento na interface. O RSVP é responsável em levar a requisição através da rede, visitando cada nodo usado para carregar o stream. Em cada nodo, o RSVP procura fazer uma reserva de recursos para o stream usando seu próprio módulo de controle de admissão, exclusivo para o RSVP, o qual determina quando o nodo tem recursos suficientes para suportar a requisição de QoS.

Se o recurso não está disponível ou o usuário não é permitido, o programa RSVP retorna uma notificação de erro para o processo da aplicação que originou a requisição. Se ocorrer sucesso, o daemon RSVP configura os parâmetros em um classificador de pacotes e scheduler de pacotes para obter o QoS desejado. O classificador de pacotes determina a classe QoS para cada pacote e o scheduler ordena a transmissão de pacotes para alcançar a promessa de QoS para cada stream. O WFQ ou WRED configura a classificação e o scheduler de pacotes requisitados para o fluxo reservado.

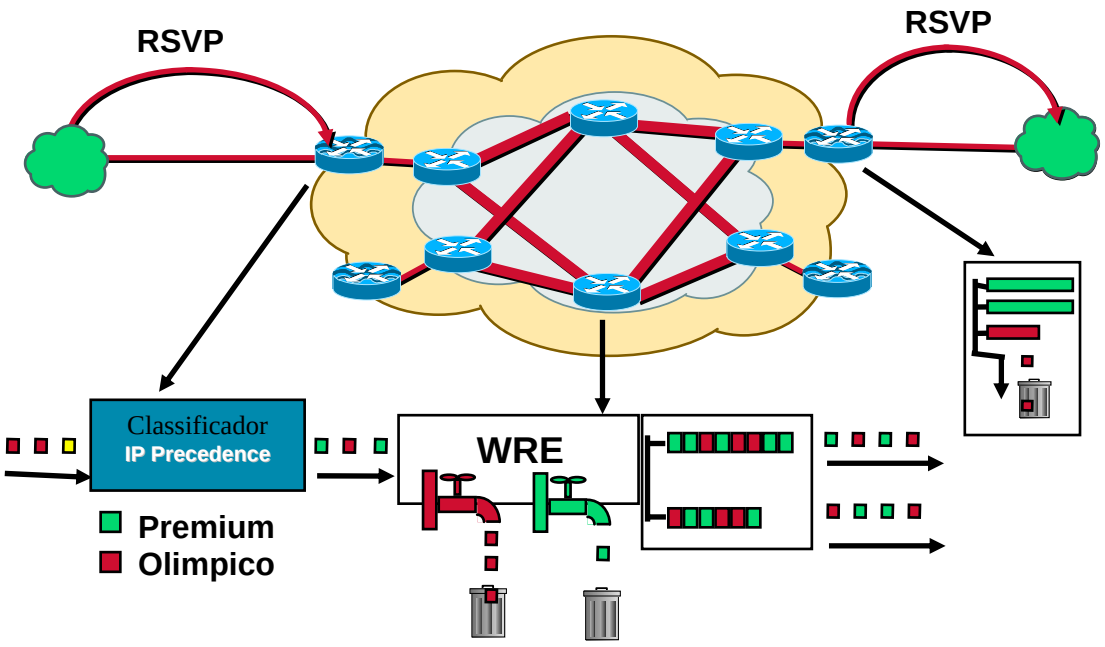

FIGURA 8.6.2 - RSVP com Weighted RED

9 A arquitetura QBone

Esta seção tem como objetivo apresentar a arquitetura necessária para participar do QBone – um testbed para Serviços Diferenciados. A idéia é especificar os requisitos mínimos necessários para que uma rede possa participar no QBone e suportar Serviços Diferenciados entre domínios [TEI99].

9.1 Introdução

O objetivo do QBone é prover uma arquitetura de teste entre domínios para Serviços Diferenciados, onde a engenharia, o comportamento e o policiamento de novos serviços IP possam ser explorados.

Cada participante da rede em um QBone é considerado um “DS domain” e a união destas redes – o próprio QBone – um “DS Region”. Os participantes devem cooperar para prover um ou mais serviços entre domínios, além do tradicional modelo de serviço “best effort”. O primeiro serviço a ser disponibilizado é o Virtual Leased Line. Todo o DS domain deve suportar um per-hop behavior (PHB) do tipo expedited forwarding (EF) e deve estar configurado para classificação e condicionamento de tráfego para prover o serviço VLL para agregados EF.

Em soma, o QBone deve suportar uma infra-estrutura integrada de medição para debugging fim a fim e auditoria de usuários, operação de rede e implementações. Tanto medições ativas como passivas devem ser coletadas e compartilhadas abertamente entre os participantes.

Outra área na qual a arquitetura QBone agrega valor à arquitetura DiffServ é em relação a operação entre domínios. No mínimo, deve ser especificado um conjunto comum de operações e procedimentos para ser seguido pelos operadores da rede. Com o progresso do QBone, existirá cada vez mais a necessidade de que operadores de rede possuam ferramentas automáticas para proceder a decisões de controle de admissão e configuração de dipositivos de rede.

9.2 Requisitos de alto-nível

A arquitetura QBone é baseada na arquitetura DiffServ; assim, ela especifica um subconjunto da arquitetura DiffServ, mais infra-estrutura de medição integrada e procedimentos operacionais pré-padronizados para estabelecer reserva entre domínios.

Os requisitos da arquitetura do QBone é contígua. Diferente do IPv6 e da tecnologia Mbone, a qualidade de serviço não pode ser implementada apenas como sendo serviços de

rede baseados na agregação de endereços. Dentro do QBone, cada participante da rede é um DS domain que interopera com outras redes QBone, para prover serviços QBone fim a fim.

Cada rede QBone deve possuir um limite administrativo bem definido, através do qual ela possa pontificar com DS domains de vizinhos.

Níveis de serviços bilaterais (SLSes) existem entre domínios DS adjacentes. Estes SLSes especificam como o tráfego é classificado, policiado e repassado pelos nodos do limite do DS.

9.3 Especificação dos serviços

9.3.1 O QBone Premium Service

O QBone Premium Service (QPS) é um serviço que garante um pico limite de largura de banda de forma simples. Sua reserva pode acontecer dentro do domínio QBone, de um limite do domínio QBone para outro limite do mesmo domínio, ou através de múltiplos domínios, definido pelo serviço-fonte e serviço-destino, cada um dos quais, em geral, define um prefixo de rede [TEI99].

Os serviços como o QPS são construídos sobre o PHB do tipo EF, de modo que a taxa de chegada dos pacotes EF em qualquer nodo é sempre menor que a taxa mínima configurada no nodo de partida, para qualquer intervalo de tempo, igual ou maior que o tempo de envio de um pacote no pico configurado.

De um modo geral, uma reserva com parâmetros QoS do tipo {source, dest, route, startTime, endTime, peakRate, MTU, jitter} provê uma transmissão assegurada do serviço QBone Premium Service.

A transmissão assegurada oferecida pelo QBone Premium Service segue os seguintes parâmetros:

- Pouca perda – não deve existir perda de pacotes; em particular, não deve existir perda de pacotes devido a congestionamentos;

- Pouca latência – o atraso dentro da fila em uma reserva QPS deve ser mínimo; entretanto, não há nada referente ao mínimo de latência no roteamento;

- Pouco jitter – o Instantaneous Delay Variation (IPDV) dentro de uma reserva QPS deve ser mínima.

Para que o serviço QPS possa assegurar a reserva, o tráfego entrante no EF deve estar conforme com o token bucket profile {peakRate, MTU}. Isso pode ser alcançado pelo

descarte dos pacotes não conformes ou pela reconformação dos pacotes até que estejam de acordo.

O parâmetro jitter é o pior caso associado à variação instantânea de atraso do pacote, ou IPDV. A medição do IPDV deve ocorrer primeiro, devido à sincronização entre os EFs ao longo do caminho do QPS. Para decidir-se sobre o controle de admissão apropriado para uma requisição de reserva QPS, é importante que um domínio Qbone entenda a variação de atraso no pior caso que um pacote EF possa experimentar.

9.3.2 Requisitos mínimos para um QBone SLS

De modo a estar consistente com a arquitetura DiffServ, todas as especificações de nível de serviço (SLSes) são determinadas bilateralmente entre os domínios DS adjacentes. Entretanto, para implementar o QBone Premium Service, os requisitos mínimos necessários para um QBone SLS devem ser [TEI99]:

(Assumindo-se um SLS bilateral entre um upstream QBone DS domain U e um downstream QBone DS domain D).

1. Dentro de um QBone, o DS Codepoint 101110 deve ser usado para o EF PHB;
2. O “D” deve responder ao pedido de reserva de um “U”;
3. Uma parte necessária de um SLS é o Traffic Conditioning Specification (TCS), que especifica quanto tráfego é condicionado e policiado no ingresso. O TCS é um componente dinâmico do SLS, o qual pode necessitar de um ajuste com a criação ou deleção de toda a reserva nos limites demarcados. Para implementar QPS, um TCS deve especificar:

    a) Condicionamento do tráfego
        1. Primeiro, o tráfego entrante deve ser condicionado no EF;
        2. Então, o tráfego EF pode ser condicionado dentro de um único EF Behavior Aggregate (BA) ou um conjunto de EF behavior aggregates, cada um dos quais pode ser definido por um prefixo-destino ou um link de saída no domínio “D”, ou outro critério.

    a) Perfis de tráfego
    Um perfile de tráfego deve ser especificado para cada behavior aggregate.
    Dado um peak rate e um MTU, o profile de tráfego é definido por um token bucket filter com:
        1. um token rate igual ao peakRate em bytes por segundo;
        2. uma profundidade de bucjet igual ao MTU em bytes.

    a) Disposição do tráfego excedente
    O tráfego dentro de um EF BA que excede o perfil do agregado deve ser descartado.

b) Conformação

A conformação dos fluxos de tráfego individuais ou agregados deve ser suportada pelos nodos da borda do Qbone, tanto na entrada como na saída.

1. O tráfego entrante no EF conforme ao profile de tráfego de um TCS terá tratamento EF através do domínio DS “U” até o destino;
2. Todo o SLS deve especificar um jitter assegurado, de modo a estar conforme ao tráfego EF.

9.4 Arquitetura de medição

Para realizar debug, auditoria e estudo dos serviços QBone, um conjunto de requisitos de medição de QoS está integrado à arquitetura do QBone. O projeto inicial do QBone Mea bsurement Architecture (QMA) está focado na verificação dos objetivos do serviço QPS, na ajuda com o debugging, provisionamento e entendimento do EF behavior aggregates. Os objetivos iniciais do QPS são: variação de perda e atraso, então medi-los passa a ser prioridade. Para ajudar com o provisionamento e entender o uso do EF SLS, a quantidade de reserva de banda relativa para a carga instantânea dos EF na entrada e saída dos links deverá ser medida. Para ajudar com o debugging e falhas isoladas, medições são necessárias para medir domínio a domínio, com pontos de medidas nas interfaces entre domínios. É recomendado que medidas fim a fim sigam o mesmo processo de medição [TEI99].

10 Qualidade de Serviço na prática

Uma vez que já foram definidas as técnicas para empregar QoS, como também a definição da arquitetura DiffServ, nesse capítulo será apresentada uma solução de Qualidade de Serviço. Nela o ambiente em questão requer diferenciação de tráfego, devido ao fato da necessidade de cobrança diferenciada do mesmo, bem como garantia do Service Level Agreement.

10.1 Topologia de rede

A topologia de rede da solução é dividida em três níveis.

- Acesso
- Distribuição
- Core

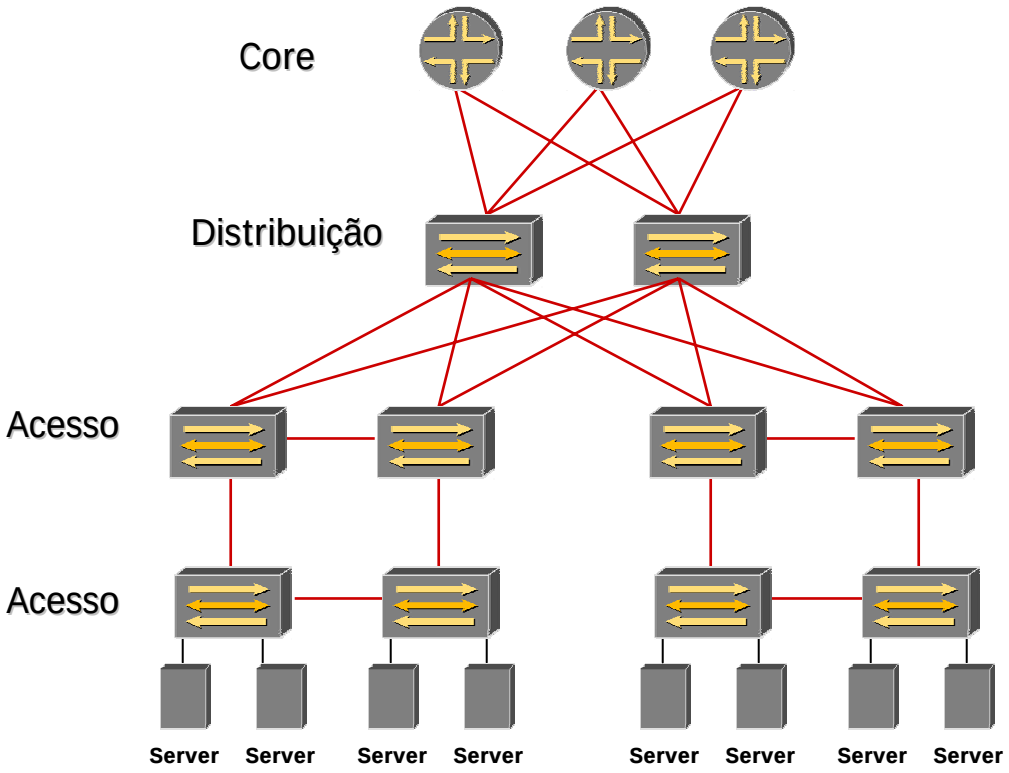

FIGURA 10.1 - Níveis de rede

10.2 Acesso

É no nível de acesso que o policiamento deve ser empregado. Isso ocorre devido ao fato da necessidade de limitar a taxa do tráfego de cada cliente requerido pelo SLA. Ou seja, cada cliente dentro da solução é cobrado pela banda que utiliza. Assim, para cada cliente o policiamento deve ser empregado.

Nesse nível também é classificado o tráfego do cliente de modo a se enquadrar em um dos seguintes tipos (classificação por DiffServ):

- Gold traffic
- Silver traffic
- Bronze traffic

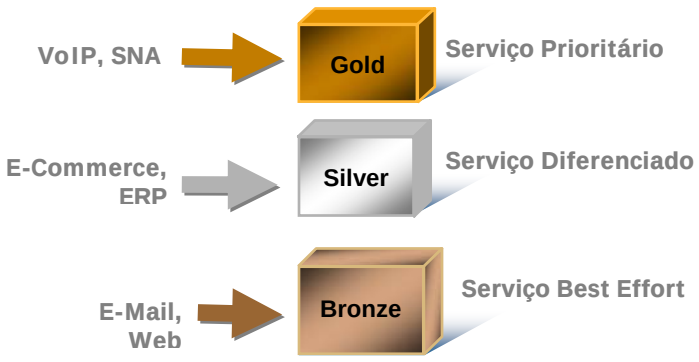

FIGURA 10.2 - Serviços diferenciados

Técnicas para garantir o SLA do cliente

Commited Access Rate (CAR)

- Marcação / Classificação de pacotes – IP Precedence e QoS group setting
- Gerência de acesso à banda – limitação de banda (policiamento)

Class-Based WFQ

- Configuração de limite mínimo de bandwidth
- Serviço de enfileiramento para controlar a latência

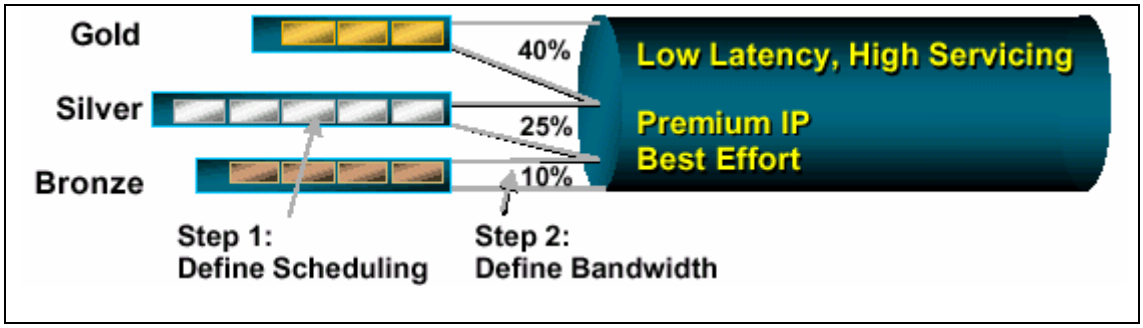

FIGURA 10.3 - Class-Based WFQ

10.3 Distribuição

O nível de distribuição tem como função definir quais serão as políticas de descarte de rede quando a rede estiver congestionada. Ou seja, a distribuição tem como função evitar e gerenciar o congestionamento da rede. Assim, quando ocorre um congestionamento o tráfego Silver ou Bronze é descartado, de modo a não afetar o tráfego Gold.

O que vale ressaltar é que o tráfego que será descartado em caso de congestionamento será o tráfego de um agregado de menor priorização e não aleatório.

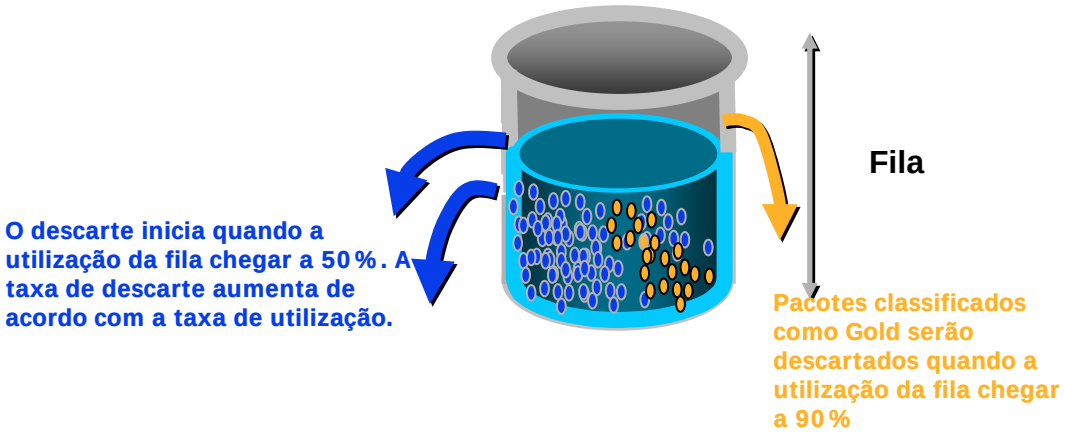

FIGURA 10.4 - Técnica de descarte

Técnicas para evitar e gerenciar congestionamento

Random Early Detect (RED) -

- Sem RED, quando a fila lota, todos os pacotes que chegam são descartados.
- Com RED, em oposto ao tail drop, o roteador monitora o tamanho médio da fila e randomicamente descarta conexões para que o congestionamento não persista.

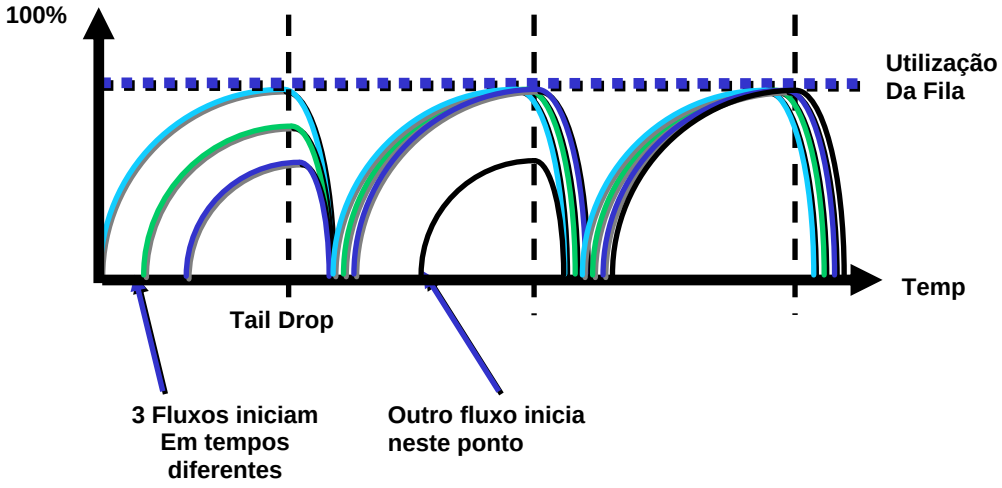

FIGURA 10.5 - Sincronismo Global

10.4 Core

Como o nível de distribuição, o Core também tem a função de prevenir o congestionamento da rede. Esse nível tem também a responsabilidade de interagir com o tráfego de ingresso e de saída. Ou seja, é nesse nível que os acordos bilaterais para fazer valer o QoS entre domínios devem ocorrer.

Técnicas para evitar e gerenciar congestionamento

Random Early Detect (RED) -

- Sem RED, quando a fila lota, todos os pacotes que chegam são descartados.
- Com RED, em oposto ao tail drop, o roteador monitora o tamanho médio da fila e randomicamente descarta conexões para que o congestionamento não persista.

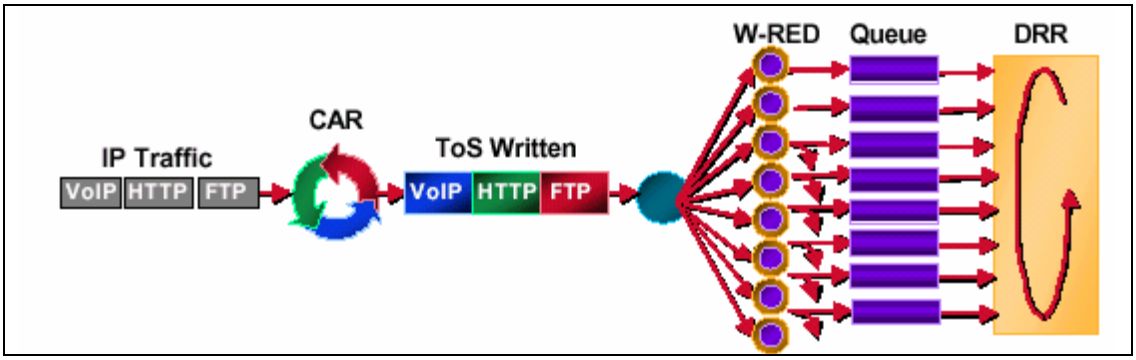

FIGURA 10.6 - Nível de Core

Assim, nesse nível os pacotes podem ser classificados e potencialmente descartados pelo W-RED (gerenciamento de congestionamento), podem ser assinalados a uma fila de saída apropriada, a transmissão pode ser programada e os pacotes podem ser re-marcados na entrada.

11 Conclusão

A Internet global está mudando tudo e todos. O mundo está convergindo para o protocolo da Internet e para suas necessidades de rede. Porém, nesse processo, o projeto original do IP tornou-se muito débil. Como resultado, a Internet necessita mudar para acomodar a demanda de novas aplicações. A largura de banda é uma solução, mas não é suficiente. De um modo geral, a Internet necessita gerenciamento dessa largura de banda, ela necessita de “inteligência”.

Até agora, a Internet tem provido apenas o serviço “best-effort”, no qual os recursos são compartilhados igualmente. Somar QoS significa somar “inteligência” à Internet, o que é uma necessidade imediata, uma vez que habilita a possibilidade de diferenciação dos serviços.

Nos provedores, esta necessidade torna-se uma prioridade, uma vez que precisamos disponibilizar mais formas de acomodar o mundo da Internet em um ambiente rico em serviços. Pois, com este tipo de arquitetura, os ISPs poderão segmentar seu mercado, entregar mais valor ao cliente, alcançar negócios mais lucrativos e esquentar o crescimento mundial na Internet.

A Internet2, com o Internet2 QoS Working Group, tem trabalhado com as aplicações da Internet2 para identificar os requisitos para QoS e as melhores tecnologias para empregá-lo. Embora não exista uma solução que satisfaça todos os requisitos identificados, o modelo de serviços diferenciados passa a ser o mais promissor.

Glossário

Behavior Aggregate (BA)	Agregação de comportamento (BA) de acordo com o campo DS.
BA classifier	Um classificador que seleciona pacotes baseado no conteúdo do campo DS.
Boundary link	Um link de conexão dos nodos de borda entre dois domínios.
Classificador	Uma entidade, que seleciona pacotes baseados no conteúdo do cabeçalho do pacote, de acordo com regras definidas.
DS behavior aggregate	Uma coleção de pacotes com o mesmo DS codepoint cruzando um link em uma direção particular.
DS boundary node (Nodo de borda)	Um nodo DS que conecta um domínio DS a um outro domínio DS ou a um outro domínio sem capacidades DS.
DS-capable	Capacidade de implementar serviços diferenciados; usado geralmente em referência a domínios que consistem de nodos de acordo com DS.
DS codepoint	Um valor específico da porção DSCP do campo DS, usado para selecionar um PHB.
DS-compliant	Habilitado a suportar funções e comportamentos de serviços diferenciados; usado geralmente para nodos e dispositivos.
DS domain	Um domínio DS-capable; conjunto de nodos os quais operam com um conjunto comum de policiamento de provisionamento de serviços e definições de PHB.
DS egress node	Um nodo localizado no limite do domínio, responsável em manipular o tráfego no momento que o mesmo deixa o domínio DS.

DS ingress node	Um nodo localizado no limite do domínio, responsável em manipular o tráfego no momento que o mesmo entra no domínio DS.
DS interior node	Um nodo DS que não é um DS boundary node.
DS field	O byte TOS no cabeçalho IPv4 ou o octeto Traffic Class no IPv6.
DS node	Um nodo DS-compliant.
DS region	Um conjunto contíguo de domínios DS os quais podem oferecer serviços diferenciados através de caminhos ao longo destes domínios.
Downstream DS Domain	O fluxo de tráfego que deixa o domínio por um link de borda.
Descartador	Um dispositivo que realiza descarte.
Dropping (Descarte)	O processo de descarte de pacotes, baseado em regras específicas; policiamento.
Legacy node	Um nodo que implementa IPv4 Precedence igual ao definido na RFC 791 e RFC 1812, mas não é DS-compliant.
Marcador	Um dispositivo que realiza marcação.
Marcação	O processo de setar o DS codepoint em um pacote baseado em regras definidas; pré-marcação, remarcação.
Mechanism	Um algoritmo específico que é implementado em um nodo para realizar um conjunto de um ou mais per-hop behaviors.
Medidor	Um dispositivo que realiza medição.
Medição	O processo de medir propriedades temporais de um tráfego selecionado pelo classificador. O estado instantâneo deste processo pode ser usado para afetar a operação de um marcador, conformador, descartador ou pode

	ser usado para o propósito de accounting e measurement.
Microflow	Uma única instância de um fluxo de pacotes de aplicação a aplicação, a qual é identificada pelo endereço origem, porta origem, endereço destino, porta destino e identificação do protocolo.
MF Classifier	Um classificador do tipo multi-field (MF) que seleciona pacotes baseado no conteúdo de alguns campos arbitrários do cabeçalho; geralmente algumas combinações de endereçamento de origem, endereçamento de destino, campo DS, protocol ID, porta origem e porta destino.
Per-Hop-Behavior (PHB)	O comportamento de encaminhamento aplicado em nodos DS-compliant para um DS behavior aggregate.
PHB group	Um conjunto de um ou mais PHBs que podem apenas ser especificados e implementados simultaneamente, devido a limites aplicados a todos os PHBs em um conjunto, tal como: um serviço de filas ou policiamento de gerência de filas. Um grupo PHB provê um service building block, que permite um conjunto de comportamento de encaminhamento para ser juntamente especificado. Um único PHB é um caso especial de um grupo PHB.
Policiamento	O processo de descarte de pacotes (por um dropper) dentro de uma seqüência de tráfego, de acordo com o estado de uma medição correspondente a um perfil de tráfego.
Pre-mark	Seta um DS codepoint de um pacote antes de entrar dentro de um downstream DS domain.
Provider DS domain	Um provedor DS-capable de serviços para um domínio origem.
Re-mark	Muda o DS codepoint de um pacote, geralmente através de um marcador, de acordo com um TCA.

Service	O tratamento global de um subconjunto definido de tráfego de um cliente, dentro de um domínio DS, ou fim a fim.
Service Level Agreement (SLA)	Um contrato de serviço entre um cliente e um provedor de serviço, que especifica o serviço de encaminhamento que um cliente deve receber. Um cliente deve ser um usuário, organização (domínio origem) ou outro domínio DS (upstream domain). Um SLA pode incluir regras de condicionamento de tráfego as quais constituem um TCA por completo ou em partes.
Service Provisioning Policy	Um policiamento que define como os condicionadores de tráfego são configurados em nodos do tipo DS boundary; e como seqüências de tráfego são mapeadas para um DS behavior aggregate, para alcançar um range de serviços.
Conformador	Um dispositivo que realiza conformação.
Conformação	O processo de atrasar pacotes dentro de uma seqüência de tráfego, de modo a estarem conformes ao perfil de tráfego definido.
Source domain	Um domínio que contém o nodo(s) que recebe o tráfego de um serviço em particular.
Traffic conditioner (Condicionador de tráfego)	Um entidade que realiza funções de condicionamento de tráfego; pode conter medidores, marcadores, descartadores e conformadores. Os condicionadores de tráfego são geralmente implantados apenas em nodos localizados no limite do DS. Um condicionador de tráfego pode remarcar o tráfego ou pode descartar ou conformar pacotes, alterando as características temporais do tráfego, de modo a levá-lo à complexidade com um perfil de tráfego.
Traffic conditioning (Condicionamento de tráfego)	Controla funções de performance através de regras especificadas em um TCA, incluindo medição, marcação, conformação e policiamento.

Traffic Conditioning Agreement (TCA)	Uma agregação especificando regras de um classificador, perfis de tráfego, medição, marcação, descarte e/ou conformação, as quais são aplicadas na seqüência de tráfego selecionada pelo classificador. Um TCA inclui todas as regras de condicionamento de tráfego explicitamente especificadas dentro de um SLA, juntamente com todas as regras implícitas nos requisitos de um Service Provisioning Policy em um domínio DS.
Perfil de tráfego	Uma descrição das propriedades temporais de uma seqüência de tráfego, tais como: taxa e tamanho de rajada.
Traffic stream (Seqüência de tráfego)	Um conjunto administrativamente significante de um ou mais microflows que cruzam um caminho. Uma seqüência de tráfego pode consistir de um conjunto de microflows ativos, os quais são selecionados por um classificador particular.
Upstream DS Domain	O fluxo de tráfego que entra no domínio por um link de borda.

Bibliografia

[BER99] BERNET, Y. et al. **A Framework for Differentiated Services.** Disponível em: <http://www.ietf.org/internet-drafts/draft-ieft-diffserv-framework-2.txt>. Acesso em: 1999.

[BER99a] BERNET, Y. et al. **Interoperation of RSVP/Int-Serv and Diff-Serv Networks.** Disponível em: <http://www.ietf.org/internet-drafts/draft-ietf-diffserv-rsvp-02.txt>. Acesso em: 1999.

[BER99b] BERNET, Y. et al. **Requirements of Diff-serv Boundary Routers.** Disponível em: <http://www.ietf.org/internet-drafts/draft-bernet-diffedge-01-Boundary-Routers.txt>. Acesso em: 1999.

[BLA99] BLAKE, S. et al. **An Architecture for Differentiated Services.** Disponível em: <http://www.ietf.org/rfc/rfc2475.txt>. Acesso em: 1999.

[BOR99] BORDEN, M.; QHITE, C. **Management of PHBs.** Disponível em: <http://www.ietf.org/internet-drafts/draft-ietf-diffserv-phb-mgmt-0.txt>. Acesso em: 1999.

[BYL99] BYLE, J. ; COHEN, R. **The COPS (Commom Open Policy Service) Protocol.** Disponível em: <http://www.ietf.org/internet-drafts/ Internet-draft/draft-ietf-rap-cops-05>. Acesso em: 1999.

[CIS99] CISCO. **Cisco IOS Quality of Service.** Disponível em: <http://www.cisco.com/warp/customer/732/net_enabled/qos.html>. Acesso em: 1999.

[CIS99a] CISCO. **Queuing, Traffic Shaping, and Filtering.** Disponível em: <http://www.cisco.com/warp/customer/732/net_enabled/queuing.html>. Acesso em: 1999.

[CIS99b] CISCO. **QoS Signaling.** Disponível em: <http://www.cisco.com/warp/customer/732/net_enabled/qos_signaling.html >. Acesso em: 1999.

[CIS99c] CISCO. **Qos Management Policy Control and Accounting.** Disponível em: <http://www.cisco.com/warp/customer/732/net_enabled/qos_management.html>. Acesso em: 1999.

[CIS99d] CISCO. **Advanced Traffic Management and QoS Concepts.** Disponível em: <http://www.cisco.com/networkers/presentations/>. Acesso em: 1999.

[FER99] FERRARI, T. **QoS Support for Integrated Networks.** [S.l.]: Faculty of Engennering, University of Bologna, 1999.

[FLO93] FLOYD, S.; JACOBSON, V. Random Early Detection Gateways for Congestion Avoidance. **IEEE/ACM Transactions on Networking**, Piscataway, NJ, v.1, n. 4, Aug. 1993.

[FLO99] FLOYD, S. **Ns Simulator Tests for Random Early Detection (RED) Gateways.** Disponível em: <http://www-nrg.ee.lbl.gov/nrg-papers.html>. Acesso em: 1999.

[FLO99a] FLOYD, S.; FALL, K. **Router Mechanisms to Support End-to-End Congestion Control.** Disponível em: <http://www-nrg.ee.lbl.gov/nrg-papers.html>. Acesso em: 1999.

[HEI99] HEINAMEN, J. et al. **Assured Forwarding PHB Group.** Disponível em: <http://www.ietf.org/internet-drafts/draft-ietf-diffserv-af-06.txt>. Acesso em: 1999.

[HOE99] HOE, J. **Improving the start-up Behaviors of a Congestion Control Scheme for TCP.** Disponível em: <http://www-nrg.ee.lbl.gov/ns/>. Acesso em: 1999.

[INL99] GOGAN, J. **IP Quality of Service**: IntServ and DiffServ. Disponível em: <http://www.unc.edu/~gogan/qos/>. Acesso em: 1999.

[IET99] IETF WORKING GROUP. **Differentiated Services.** Disponível em: <http://www.ietf.org/html-charters/diffserv-charter.html>. Acesso em: 1999.

[JAC99] JACOBSON, V.; NICHOLS, K. **An Expedited Forwarding PHB.** Disponível em: <http://www.ietf.org/internet-drafts/draft-ietf-diffserv-phb-ef-02.txt>. Acesso em: 1999.

[LI99] LI, T.; REKHTER, Y. **A Provider Architecture for Diferentiated Services and Traffic Enginnering.** Disponível em: <http://www.ietf.org/rfc/rfc2430.txt>. Acesso em: 1999.

[MAC99] MACKIE, J.; VARIAN, H. **Pricing the Internet.** Disponível em: <ftp://ftp.econ.lsa.umich.edu/pub/Papers/Pricing_the_Internet.ps.Z>. Acesso em: 1999.

[MEH99] MEHRA, A.; DINESH, V. **Differentiated Services Operational Model and Definitions.** Disponível em: <http://www.ietf.org/internet-drafts/draft-nichols-dsopdef-00.txt>. Acesso em: 1999.

[MEH99a] MEHRA, A.; DINESH, V. **Architectural Considerations for DiffServ Servers.** Disponível em: <http://www.ietf.org/internet-drafts/draft-mehra-diffserv-servers-00.txt>. Acesso em: 1999.

[MIL99] MILINKOVICH, M. DiffServ Migration Planning. **Cisco Systems Users Magazine,** San Jose, v. 11, p. 71-73, Oct. 1999.

[NIC99] NICHOLS, K. et al. **Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers.** Disponível em: <http://www.ietf.org/rfc/rfc2474.txt>. Acesso em: 1999.

[NIC99a] NICHOLS, K.; BLAKE, S. **Differentiated Services Operational Model and Definitions.** Disponível em: <http://www.ietf.org/internet-drafts/draft-nichols-dsopdef-00.txt>. Acesso em: 1999.

[NIC99b] NICHOLS, K. et al. **A Two-bit Differentiated Services Architecture for the Internet.** Disponível em: <http://www.ietf.org/internet-drafts/draft-nichols-diff-svc-arch-00.txt>. Acesso em: 1999.

[RFC1163] LOUGHEED, K.; REKHTER, U. **Border Gateway Protocol (BGP)**. Disponível em: <http://www.ietf.org/rfc/rfc1163.txt>. Acesso em: 1999.

[RFC 791] INTERNET Protocol: RFC 791. Disponível em: <http://www.ietf.org/rfc/rfc791.txt>. Acesso em: 1999.

[SHE99] SHENKER, S. et al. **Specification of Guaranteed Quality of Service.** Disponível em: <ftp://ftp.ietf.org/internet-drafts/draft-ietf-inteserv-guaranteed-svc-06.txt>. Acesso em: 1999.

[STE99] STEPHENSON, A. **Diffserv and MPLS:** Aquality Choice. Disponivel em: <http://www.data.com/issue/981121/quality.html>. Acesso em: 1999.

[TEI99] TEITELBAUM, B. **QBone Architecture.** Disponível em: <http://www.internet2.edu/qos/wg/papers/qbArch/06/draft-i2-qbone-arch-06.html>. Acesso em: 1999.

[TEI99a] TEITELBAUM, B. **Building a Testbed for IP Differentiated Services.** Disponível em: <http://www.internet2.edu/qbone/>. Acesso em: 1999.

[TEI99c] TEITELBAUM, B. **First Internet2 Joint Applications/ Engineering QoS Workshop.** Disponível em: <http://www.internet2.edu/>. Acesso em: 1999.

[YAV99] YAVATKAR, R. et al. **COPS Usage for Differentiated Services.** Disponível em: <http://www.ietf.org/internet-drafts/draft-bernet-diffedge-01.txt>. Acesso em: 1999.

[WRO99] WROCLAWSKI, J. **Specification of the Controlled-Load Network Element Service.** Disponível em: <http://www.ietf.org/rfc/rfc2211.txt>. Acesso em: 1999.