



# Qualidade de Serviço na Internet

Redes de Computadores II  
IFCE

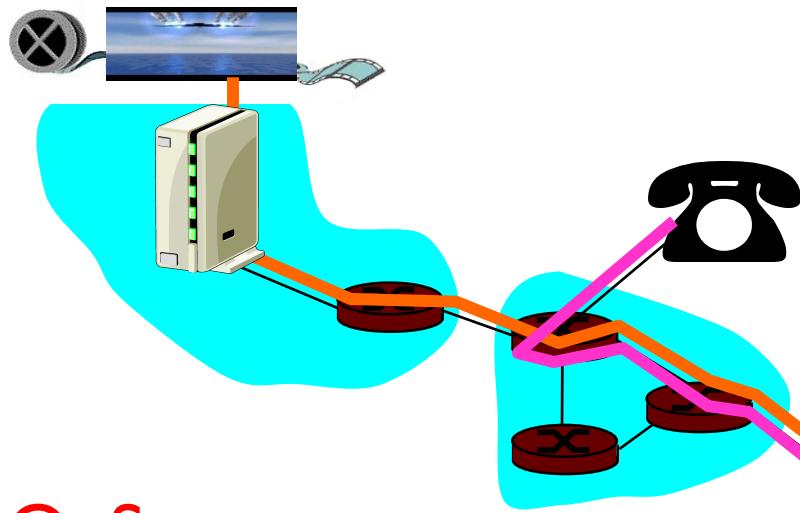
# Materiais utilizados

- Apresentação de Jim Kurose, Department of Computer Science, Univ.Massachusetts/Amherst
- Networkers da Cisco.com
- Apresentação de Paulo Carvalho e Solange Lima, DI, Univ. do Minho.
- Apresentação de Luciana Bolan Frigo e Mário Lucio Roloff, BR

# Sumário

- Introdução
- Serviços Integrados
- Problemas dos Serviços Integrados
- Serviços Diferenciados

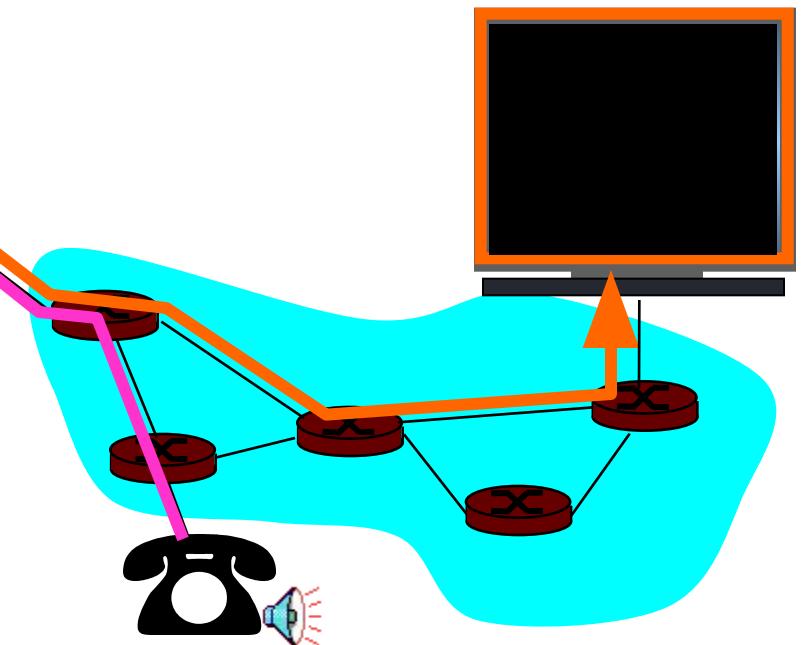
# Qualidade de Serviço: o que é?



Aplicações multimídia: Vídeo e Áudio na Internet

**QoS**

A rede disponibiliza à aplicação *o nível de desempenho necessário para o seu funcionamento*.



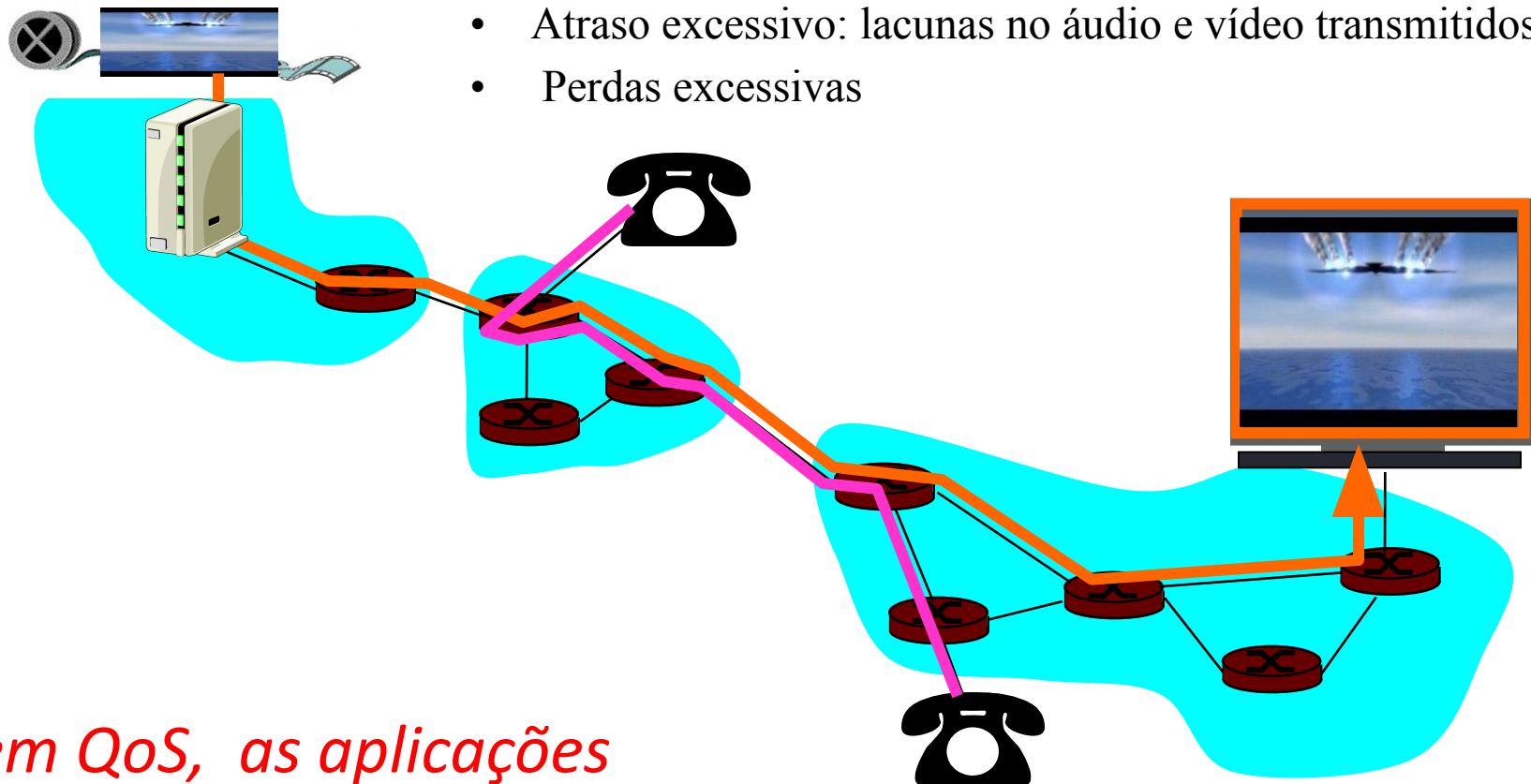
# Requisitos de QoS para as Mídias Contínuas

- Transmissões ao vivo e em demanda
- *Requisitos: entregar os dados de forma temporizada*
  - Atrasos fim-a-fim curtos para Multimídia Interativa
    - i.e., Telefone sobre IP, Teleconferência., Mundos Virtuais, Simulação Interativa Distribuída
  - Reconstituição contínua da mídia na chegada
    - Variação limitada dos atrasos dos pacotes
- Não necessidade de total confiabilidade
  - Não é sempre necessária uma de fiabilidade de 100%

# O que acontece sem QoS?

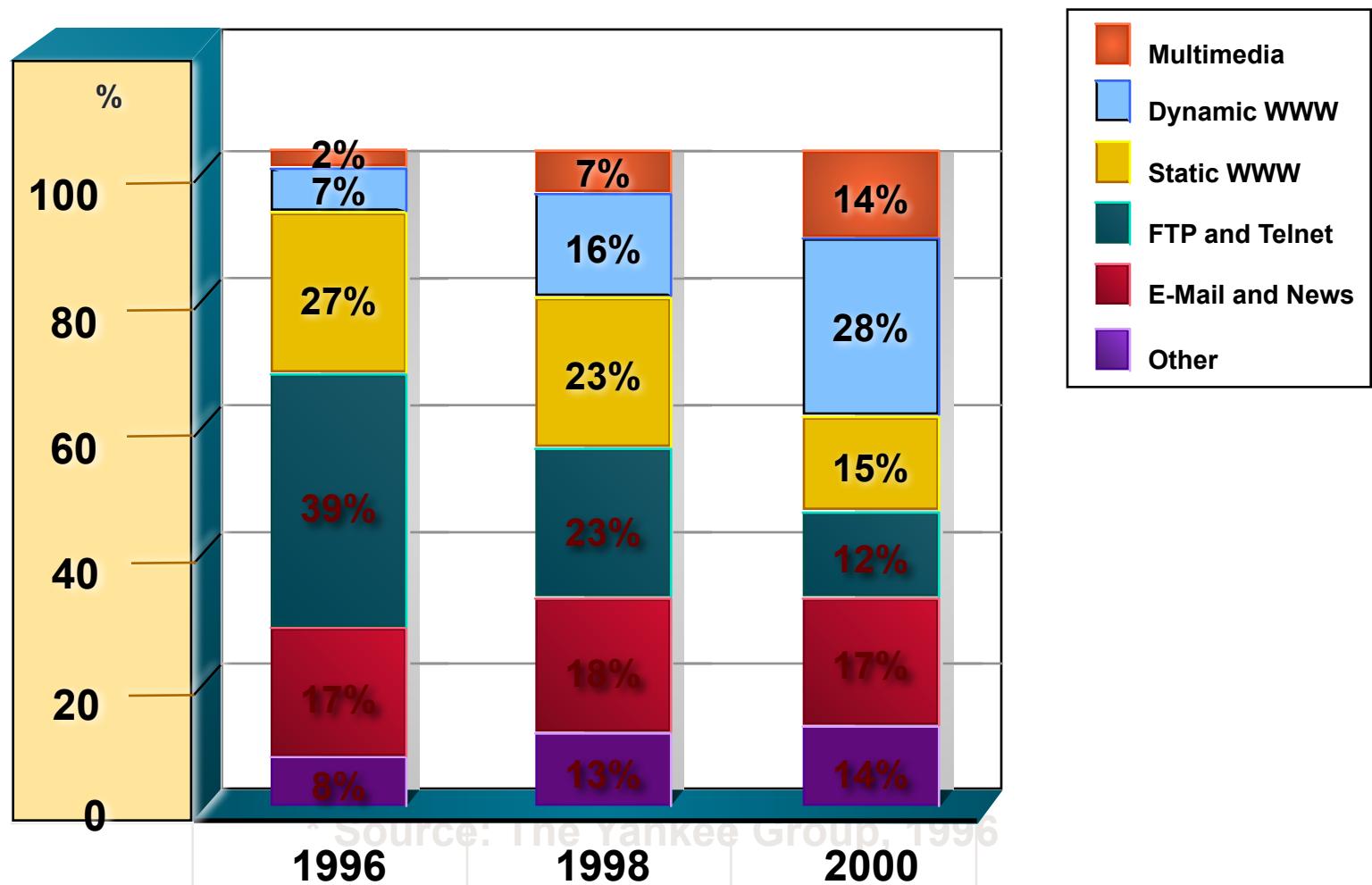
## Problemas:

- Atraso excessivo: lacunas no áudio e vídeo transmitidos
- Perdas excessivas



*Sem QoS, as aplicações  
multimídia não funcionam*

# Mudanças no tipo de tráfego de Aplicação



# Pergunta

- Há quem defende que os problemas de QoS podem ser resolvidos, aumentando a disponibilidade de largura de banda das linhas, de forma a que a limitação de largura de banda não constitua um problema. Qual é a sua opinião?

# Uma Resposta possível

- Por mais largura de banda que se disponibilize vão existir novas aplicações que a tornam sempre um recursos crítico. Além disso, o problema que se coloca não é só a falta de largura de banda mas sobretudo a forma como ela é usada. Mecanismos com a qualidade de serviço permitem a utilização e compartilhamento racional de recursos e isso deve ser considerado.

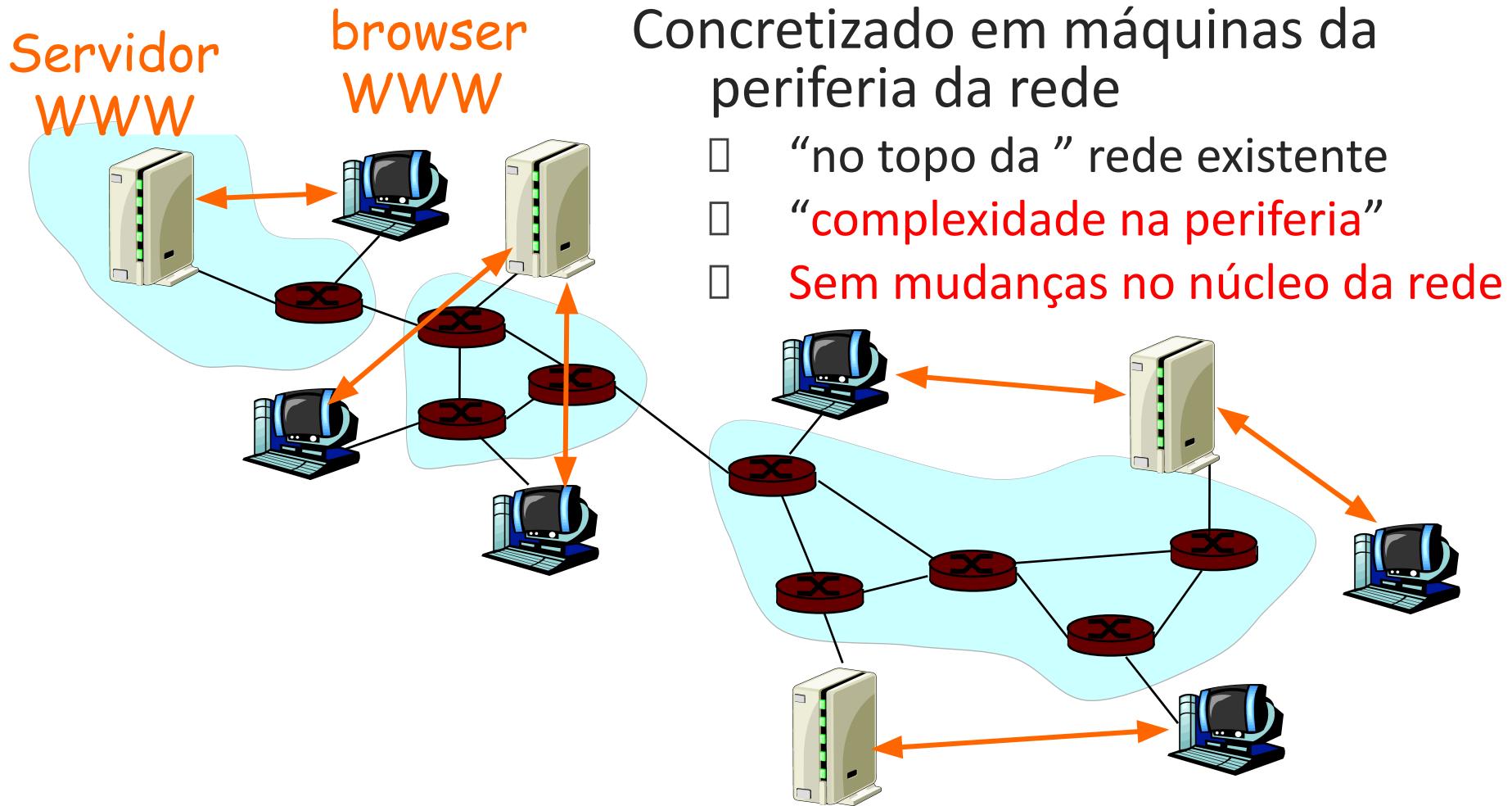
# QoS: Por que que não temos?

- QoS é uma preocupação desde sempre
  - O que aconteceu com o QoS?
  - A Internet cresceu um milhão de vezes! Aplicações: WWW, Napster, ...
- Poque há progressos pouco significativos no QoS?
    - Um monte de gente inteligente e competente trabalham nisso!

A Internet é um grande sucesso!

Por que o QoS é tão difícil?

# Por que o WWW é tão “simples”?



# Por que o QoS é difícil?

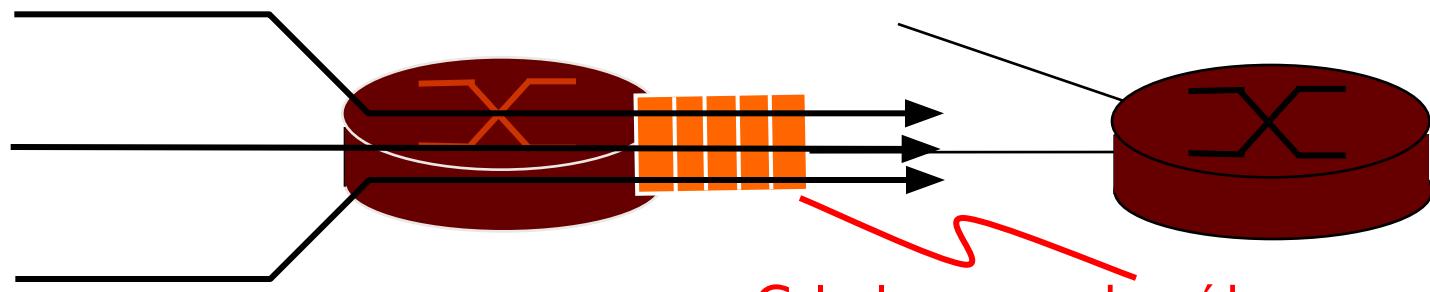
- O **núcleo da Internet** atual disponibiliza um serviço “**best effort**”
  - A congestão da rede causa atrasos, perdas
  - Sem garantia de tempos de entrega
  - Sem garantias nas perdas
- Multimídia precisa que se consigam restrições nas perdas e tempos de entrega

“As diferenças nas restrições nos tempos de entrega e confiabilidade das comunicações em tempo real requerem a concepção de novos protocolos e arquiteturas

**Até que ponto  
isto é verdade?**

# Por que o QoS é difícil?:

Para fornecer garantias de desempenho nos atrasos e perdas:



Calcular a taxa de saída na sessão

- Fornecidos o tráfego de entrada declarado e a disciplina de escalonamento

Precisa do tráfego de entrada da sessão

conhecendo o perfil de tráfego da aplicação

São necessários avanços fundamentados nesses problemas complexos

# Fluxo de Tráfego

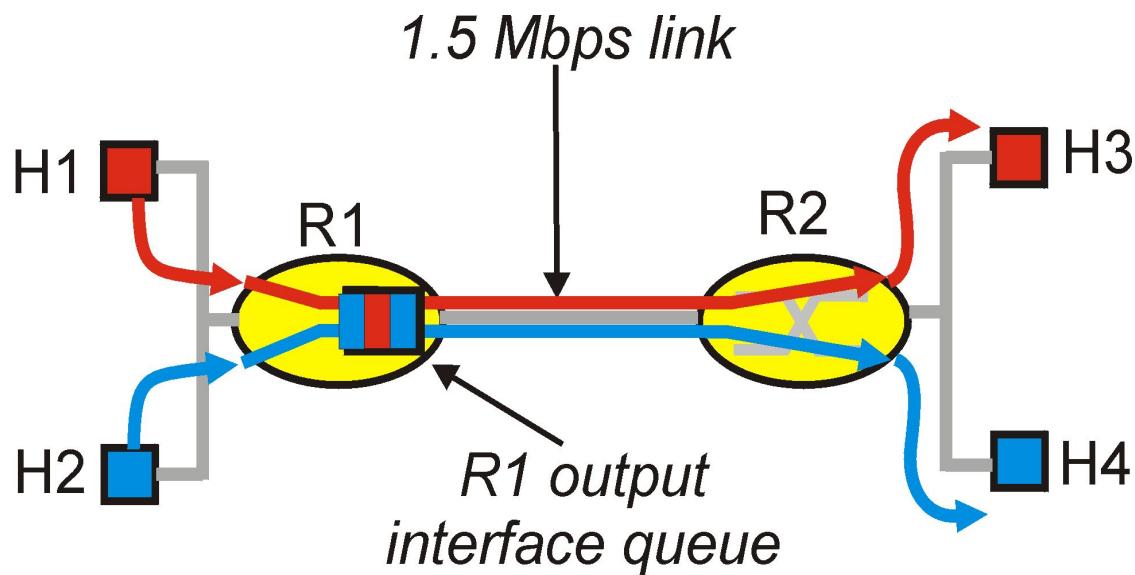
- Sequência de pacotes com:
  - Mesma máquina origem e destino
    - Endereço IP
  - Mesma aplicação origem e destino
    - Portas de Protocolo (TCP ou UDP)
  - Mesmos requisitos de QoS

# QoS vs CoS

- Qualidade de Serviço (QoS)
  - Traduz nível de desempenho pretendido
  - Expressa por um conjunto de parâmetros
    - Largura de Banda, atrasos (jitter), perdas
- Classes de Serviço (CoS)
  - Definem serviços com objectivos de QoS distintos
  - Cada classe tem um nível de QoS associado e define um tratamento a receber pelos pacotes na rede
  - Fluxos de tráfego classificados com base nos seus requisitos de QoS

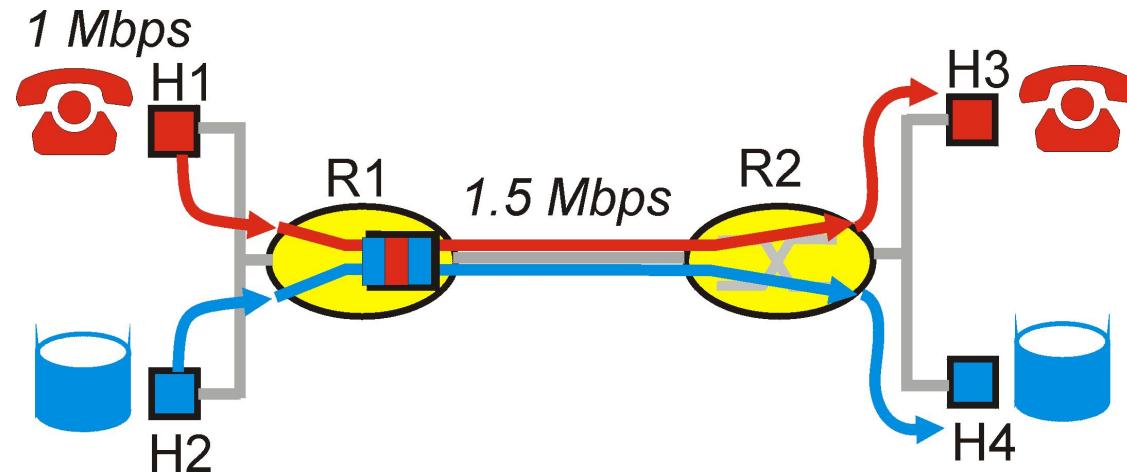
# Princípios para garantir QOS(1)

- Topologia simples para estudar compartilhamento e congestão



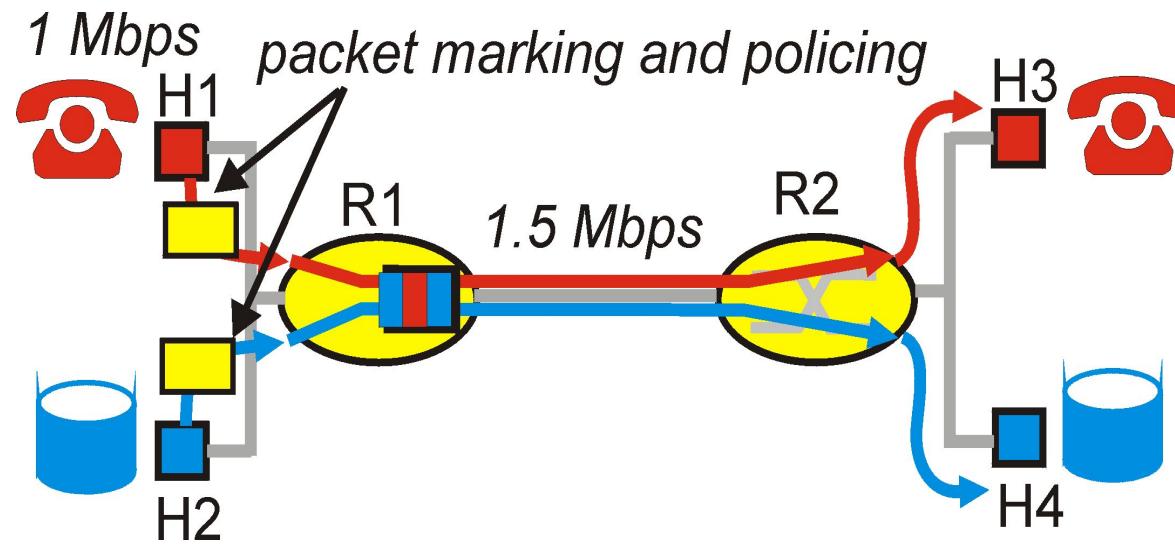
# Princípios para garantir QOS(2)

- Uma aplicação de telefone a 1Mbps e outra FTP dividem uma ligação a 1.5 Mbps.
  - Tráfego FTP pode congestionar o roteador que descarta de pacotes áudio.
  - Prioriza-se o audio relativamente ao FTP.
- **Princípio 1: é necessário marcar os pacotes para o roteador distinguir entre diferentes classes e de novas políticas para manipulação dos pacotes**



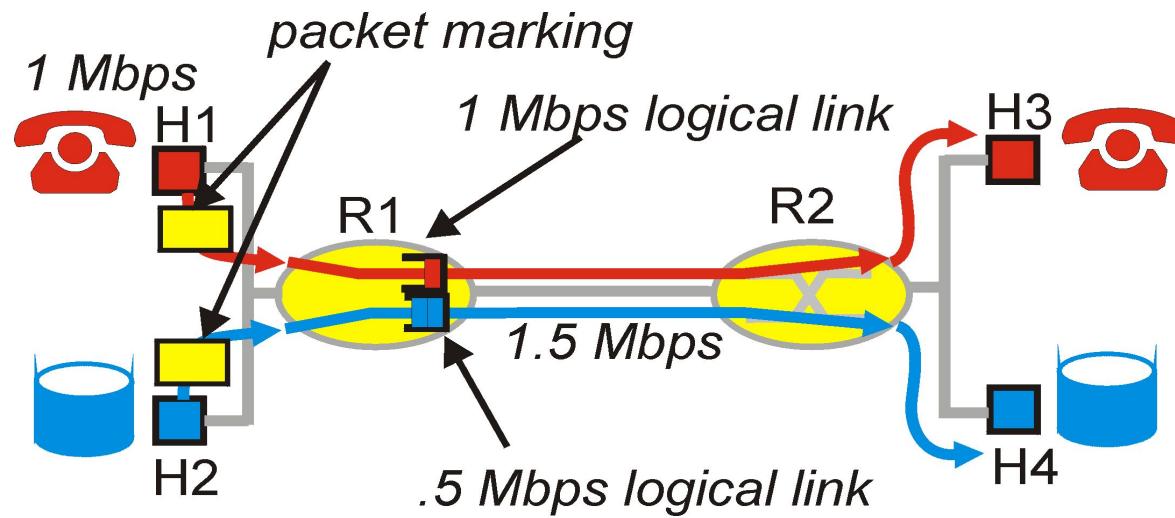
# Princípios para garantir QOS(3)

- Aplicações mal comportadas (O audio envia pacotes a um débito superior a 1 Mbps).
- Princípio 2: isolar uma classe das restantes.**
- São necessários mecanismos para policiar das fontes. A marcação e policiamento deve ser feita na periferia da rede:



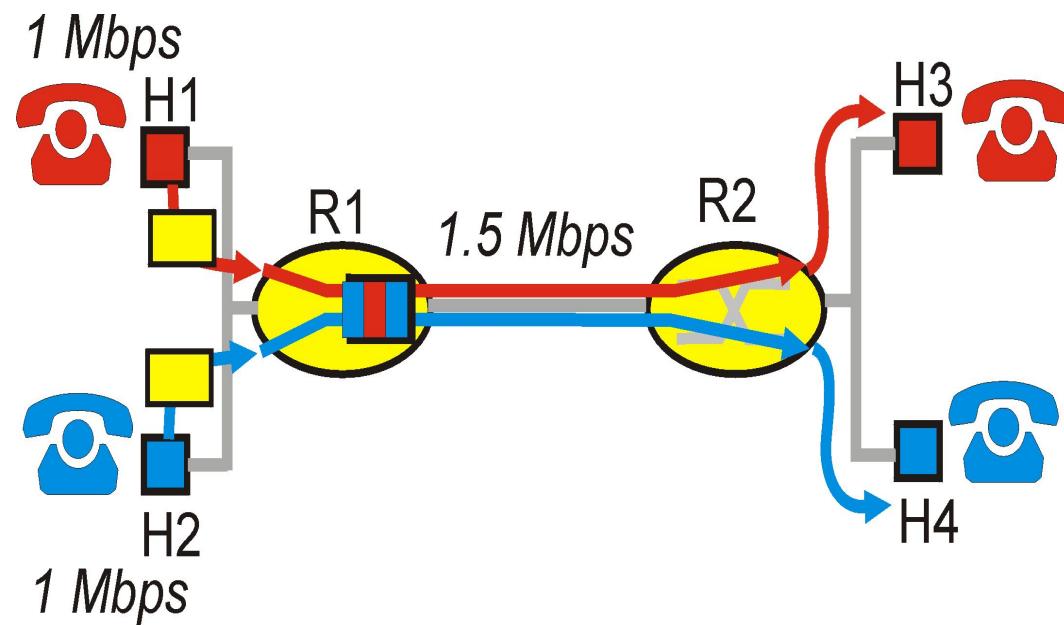
# Princípios para garantir QOS(4)

- Alternativa à marcação e ao policiamento: alocar uma porção da largura de banda à cada fluxo de aplicação; pode haver desperdício se um dos fluxos não usar a sua largura de banda
- Princípio 3: O isolamento disponibilizado deve coexistir com uma utilização eficiente dos recursos**



# Princípios para garantir QOS(5)

- Não é possível suportar tráfego além da capacidade das ligações.
- Princípio 4: É necessário um mecanismo de Controle de Admissão; o fluxo de aplicação declara as suas necessidades e a rede pode bloquear a comunicação por falta de recursos.



# Princípios para garantir QOS(6) Resumo

- QoS para aplicações de Rede
  - Classificação de pacotes
  - Isolamento
    - Escalonamento e policiamento
  - Percentagem elevada de utilização de recursos
  - Controle de Admissão

# Pergunta 1

- O que é QoS e que parâmetros se pode medir?

Qualidade de serviço pode ser definida de várias formas.

1. *diversas formas de avaliação da capacidade de rede para realizar tarefas.*
2. *oferecer um serviço de transferência adequado para satisfazer os requisitos das aplicações utilizadas.*

A qualidade de serviço em redes IP, por exemplo, pode ser medida usando os seguintes parâmetros:

- atraso na transferência,
- variação no atraso da transferência,
- erros na transferência,
- percentagem de pacotes perdidos relativamente ao total de transferidos
- e largura de banda

# Pergunta 2

□ Como é que o usuário percebe QoS?

1. disponibilidade do serviço
2. custo do serviço, já que os serviços de boa qualidade são normalmente mais caros.
3. tempo de resposta que é mais perceptível nos serviços interativos.
4. a probabilidade de falha ou de erro de serviço, uma vez garantido o seu correto funcionamento.

# Pergunta 3

- Quais são os requisitos para introdução do QoS (se existirem)?
  - serviço de rede deve ser capaz de distinguir os diversos tipos de aplicações seja por classes ou individualmente e capaz de disponibilizar recursos (largura de banda, buffers, tempo de processamento) de acordo com as restrições de cada tipo de tráfego.
  - Isto significa que os roteadores devem ser capazes de tratar os fluxos (individualmente ou por classes) de forma diferenciada. Isto implica na existência de mecanismos de reserva de recursos, controle de admissão, marcação e policiamento de tráfego.

# Pergunta 4

- Que problemas de segurança podem ser colocados através de QoS?

Alguém pode tentar “roubar” a largura de banda e os restantes recursos usados para fornecer qualidade de serviço injetando um grande número de falsos pacotes para a rede, inundando-a com tráfego.

Numa situação dessas, a Qualidade de Serviço para os usuários convencionais vai sofrer de Negação de Serviço[1] (Denial of Service).

[1] Denial of Service, na terminologia inglesa

# Modelos de Seviços Integrados (IntServ)

## *Filosofia*

- Sinalizar QoS é função da aplicação
- Garantia de QoS é obtida com reserva prévia de recursos
- Existe controle de admissão
- Roteadores reservam recursos para fluxos específicos e mantêm informação sobre seu estado (hard-state)

# Modelos de Seviços Integrados

## *Grupos do IETF*

- IntServ Working Group
  - Descrição da Arquitetura (RFC 1633)
  - Tipos de Serviço (RFC 2211 e 2212)
- RSVP Working Group
  - Descrição do Protocolo (RFC 2205)
  - RSVP no IntServ (RFC 2210)

# Serviços Integrados

## Componentes

- Protocolo de sinalização
- Reserva de recursos
- Controle de Tráfego
  - Controle de admissão
  - Classificador
  - Escalonador

# Serviços Integrados

## Protocolo de reserva de recursos

- Reserva dinâmica
- Recebe da aplicação
  - Especificação do fluxo
  - Origem e destino(s)
- Protocolo de controle
  - Não transporta dados
  - Ex: RSVP

# Serviços Integrados

## Especificação do Fluxo

- Define
  - Características do fluxo de tráfego
    - Taxa, tamanho de burst,...
  - QoS necessário ao fluxo
- Determina
  - Reserva dum conjunto de recursos mínimo
    - Largura de banda mínima
    - Identificação do tráfego em conformidade ou não com a especificação
    - Parametrização do Escalonador

# Modelo de Serviços Integrados

- Escalonador de pacotes: gere o encaminhamento dos fluxos de dados, usando alguma política de filas
- Classificador: coloca os pacotes que chegam em determinadas filas
- Controle de admissão: implementa o algoritmo p/ verificar se um novo fluxo pode ter seu pedido de QoS atendido
- Protocolo de reserva de recurso

# Síntese de operação do RSVP

- Mensagem PATH especifica característica do tráfego
- Encaminhadores do percurso expedem a mensagem e criam estado
- Mensagem de RESV requisita recursos para o fluxo
- Cada encaminhador do percurso
  - Aceita o pedido de RESV reserva LB e buffer
    - Regista informação do estado do fluxo
  - Rejeita o pedido de RESV
    - Envia mensagem de erro e termina processo de sinalização

# Síntese de Operação do RSVP



PATH  
→

PATH  
→

PATH  
→

PATH  
→



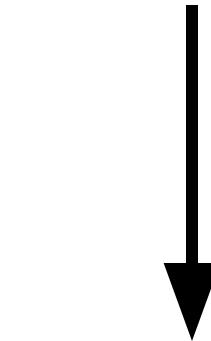
Largura de banda = X  
Atraso máx = Y

Solicitação ok

O Roteador recebe a  
mensagem PATH e  
verifica se pode  
atendê-la.

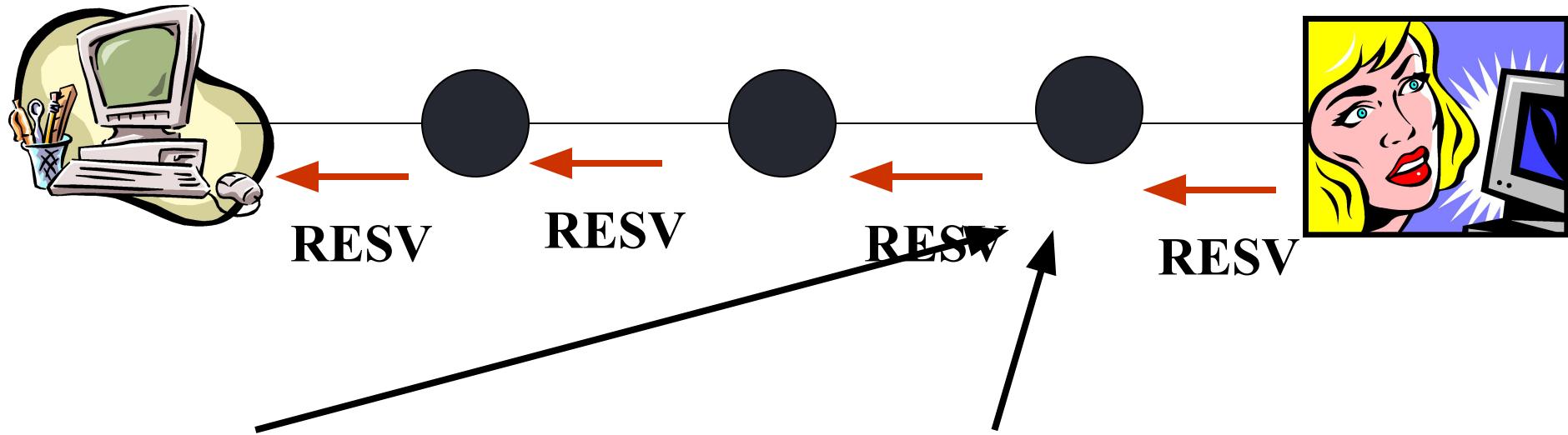
PATH

Informar as características de tráfego da  
requisição do transmissor e sobre o caminho  
fim a fim entre eles.



Todos Roteadores  
suportam RSVP

# Síntese de Operação do RSVP



**Usa o processo de controle de admissão para autenticar a solicitação e alojar os recursos necessários**

**O Roteador recebe a mensagem RESV**

**RESV**

**Faz a reserva de recurso de acordo com os parâmetros da PATH.**

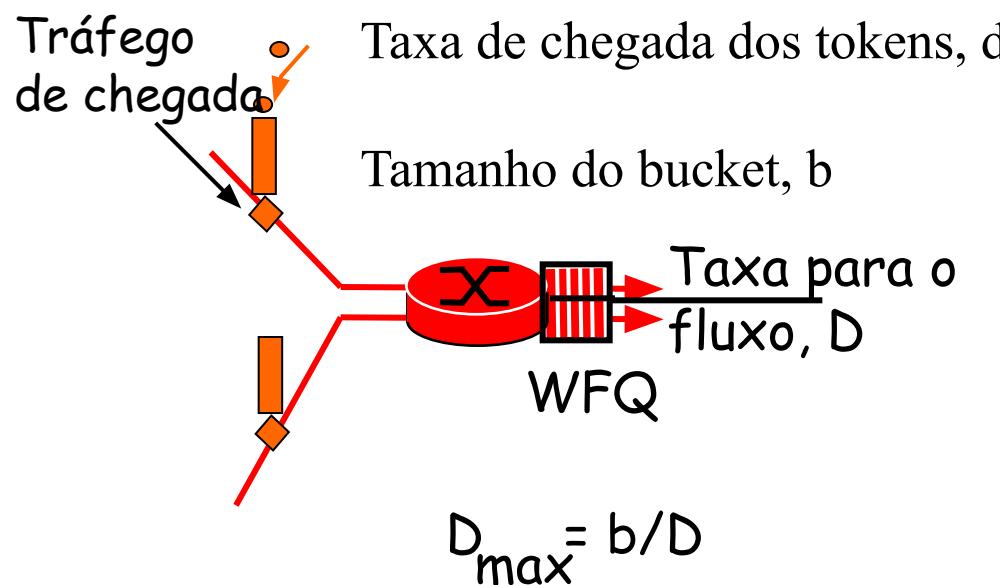
# QoS nos Serviços Integrados: Modelos de Serviço [RFC 2211, RFC 2212]

## Serviço garantido:

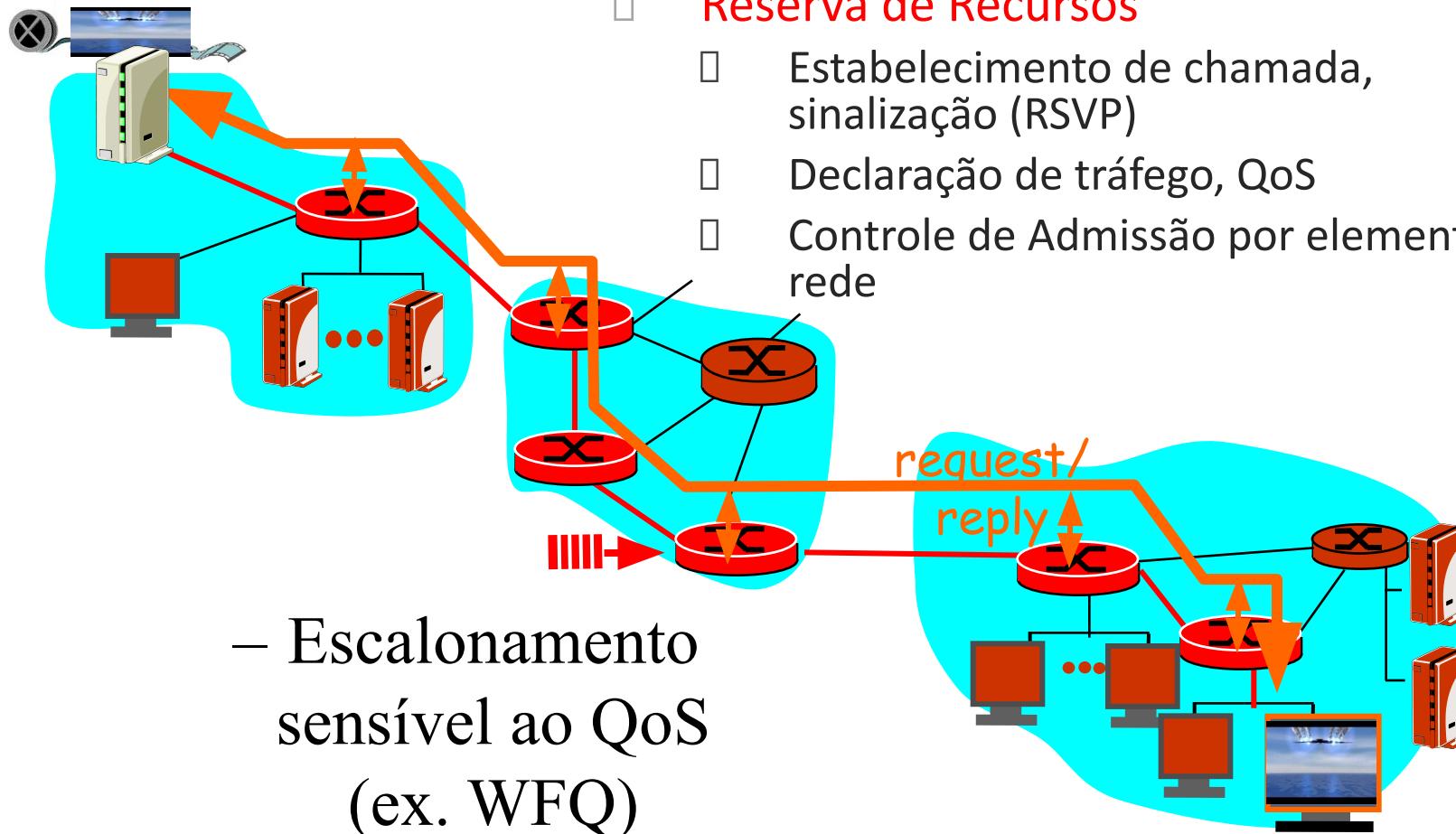
- Pior caso na chegada de tráfego: leaky-bucket-policed source
- **Limite simples** (prova-se matematicamente) *no atraso* [Parekh 1992, Cruz 1988]

## Serviço de carga controlada:

- “Qualidade de serviço bastante aproximada da QoS oferecida ao mesmo fluxo por um elemento de rede sem carga.”



# Serviços Integrados: garantias de QoS



# 1995/6: Todas peças do QoS no lugar

- Modelos de serviço
- Teoria necessária: escalonamento, política, desempenho
- Protocolo de sinalização
- Implementação de fabricantes

Últimos anos: Pouco progresso relativo  
Por que?

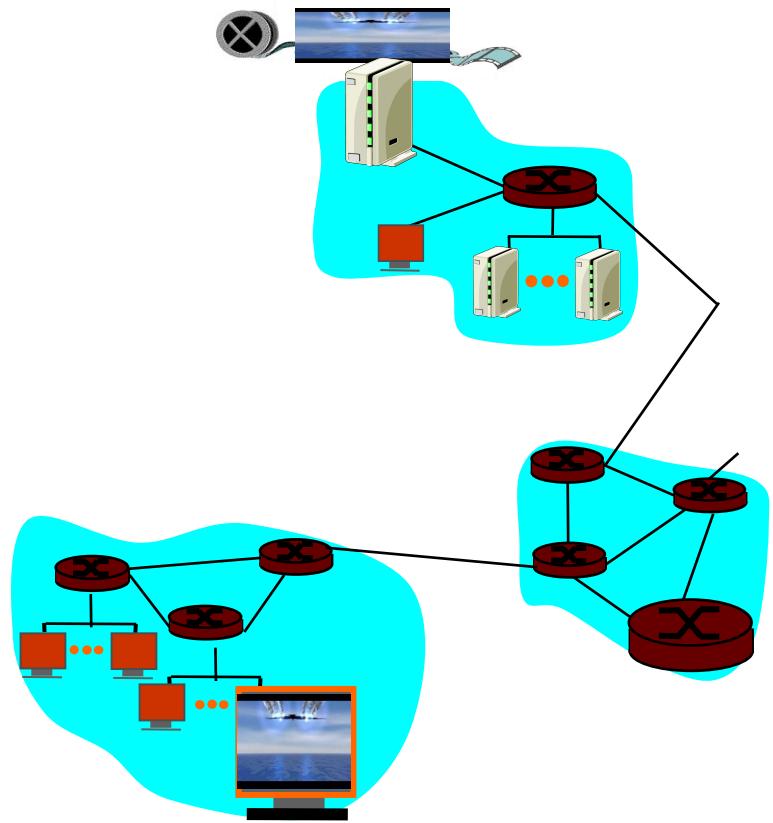


# Lição nº1: Escalabilidade das soluções

- Soluções demonstradas a pequena escala podem não funcionar em grande escala:
  - Sinalização por chamada: complexa?
    - Pouco problemática em pequenas redes
    - Um router numa rede de *backbone* modesto vê 250K fluxos/min
- **Desafio:** disponibilizar QoS à larga escala
  - Cada sessão usa múltiplos elementos
    - Cada elemento manipula um grande nº de sessões
      - Micro-análise detalhada não é possível
      - Garantias estatísticas sobre o agregado?

## Lição nº2: Introduzir nova tecnologia no núcleo da rede é difícil !

- Dificuldades similares com o IPv6 e multicast
- Mudanças devem ser incrementais
  - Problemas legais
  - Base existente muito grande-> só com uma vitória por KO da nova tecnologia relativamente a todas as outras...



# Lição nº3: QoS via A alocação

Solução: resolver (minimizar) o problema com o sobre-dimensionamento de recursos

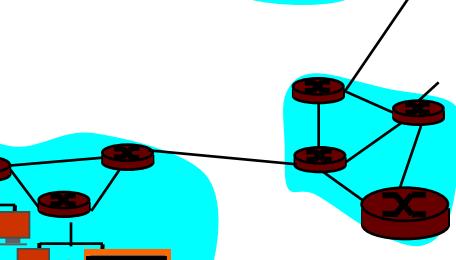
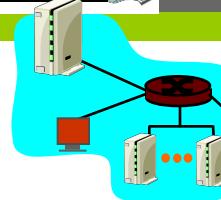
- Rede com “sobre-alocação”: torna extremamente raros grandes atrasos e perdas (perdas de QoS)
- Usada atualmente por: Sprint, AT&T
  - Rede Sprint nos EUA:  $\text{RTT} < 70 \text{ ms}$ ;  $\text{loss} < 0.01\%$
  - Outras razões para sobre-alocação:
    - Uso da largura de banda cresce exponencialmente: ISP 6 meses à frente da curva
    - Custos altos para mudar a tarifa fixa do best-effort
  - **Efeito secundário(?)**: Utilização dos recursos longe do ótimo

# Lição nº3: QoS via Alocação

## Desafios para investigação:

- Dimensionar e monitorar QoS via sobre-alocação
- Como realocar recursos com rapidez
- Comparar os custos da sobre-alocação versus QoS à medida
- Adaptatividade: simular os efeitos das falhas de QoS

# Lição nº4: A qualidade dum percurso é determinada pela pior ligação



## Internet = rede de redes

- Uma solução fim a fim exige a cooperação de *todas* redes do percurso: *difícil de conseguir!*
- Uma única rede congestionada significa falha num percurso fim-a-fim

• **Questão:** O conceito rede de redes é sinônimo de **engarrafamento**?

- Permitiu um grande crescimento e penetração da Internet
- Vantagem competitiva (fornecimento de QoS) quando há controle administrativo de todo o percurso.

# Serviços Diferenciados (DiffServ)

## Filosofia

- Definição de um pequeno número de classes de serviço
  - Com requisitos de QoS associados
  - Com filas de espera diferenciadas
- Marcação e Classificação de tráfego em classes de serviço
- Tratamento diferenciado dos pacotes
  - De acordo com a classe

# Serviços Diferenciados

## Grupos de Trabalho

- IETF
  - Descrição da arquitetura –RFC 2475
  - Campo DS –RFC 2774
  - PHBs – RFC 2397 2598
  - Internet Drafts
- TF-Tant
  - Ambiente experimental de teste de arquitetura e de serviços

# Serviços Diferenciados

- QoS na Internet para agregações de fluxos
  - Sem estado para cada fluxo de dados
  - Sem sinalização para cada nó
- DS-Field
  - Pacotes são marcados para receber serviços diferenciados nos **Domínios DS**
  - Campo TOS do IPv4 ou Traffic Class do IPv6
  - Identifica o **PHB** (Per-Hop Behavior)
  - Valores do DS-Field são chamados de **DSCP** (DiffServ Code Point)

# Contratos e Serviços

- SLA: Service Level Agreement
  - Contrato de serviço bilateral
  - Identifica perfil de tráfego (ex.  $r = 1 \text{ Mbps}$ ,  $b = 100 \text{ Kb}$ )
- Serviço: “tratamento global de um determinado subconjunto do tráfego de um utilizador dentro de um Domínio DS, ou fim a fim”
- PHBs + Policy Rules = vários serviços
- O grupo de trabalho da IETF não vai normalizar os serviços fim a fim (para os utilizadores)

# Contratos e serviços

- Para obter serviços diferenciados
  - Cliente pode ser outro ISP
  - Aspectos técnicos do contrato
    - Especificação de Nível de Serviço (Service Level Agreement –SLA)
    - Especifica
      - Serviços pretendidos
      - Perfil de tráfego
      - Propriedades temporais do tráfego
      - Regras de condicionamento de tráfego

# Contrato de Serviço

- Pode ser estático ou dinâmico
- Determina no ISP
  - Alocação do espaço em *buffers*
  - Alocação de Largura de banda
  - Como as várias aplicações partilham o SLA é da responsabilidade da fonte

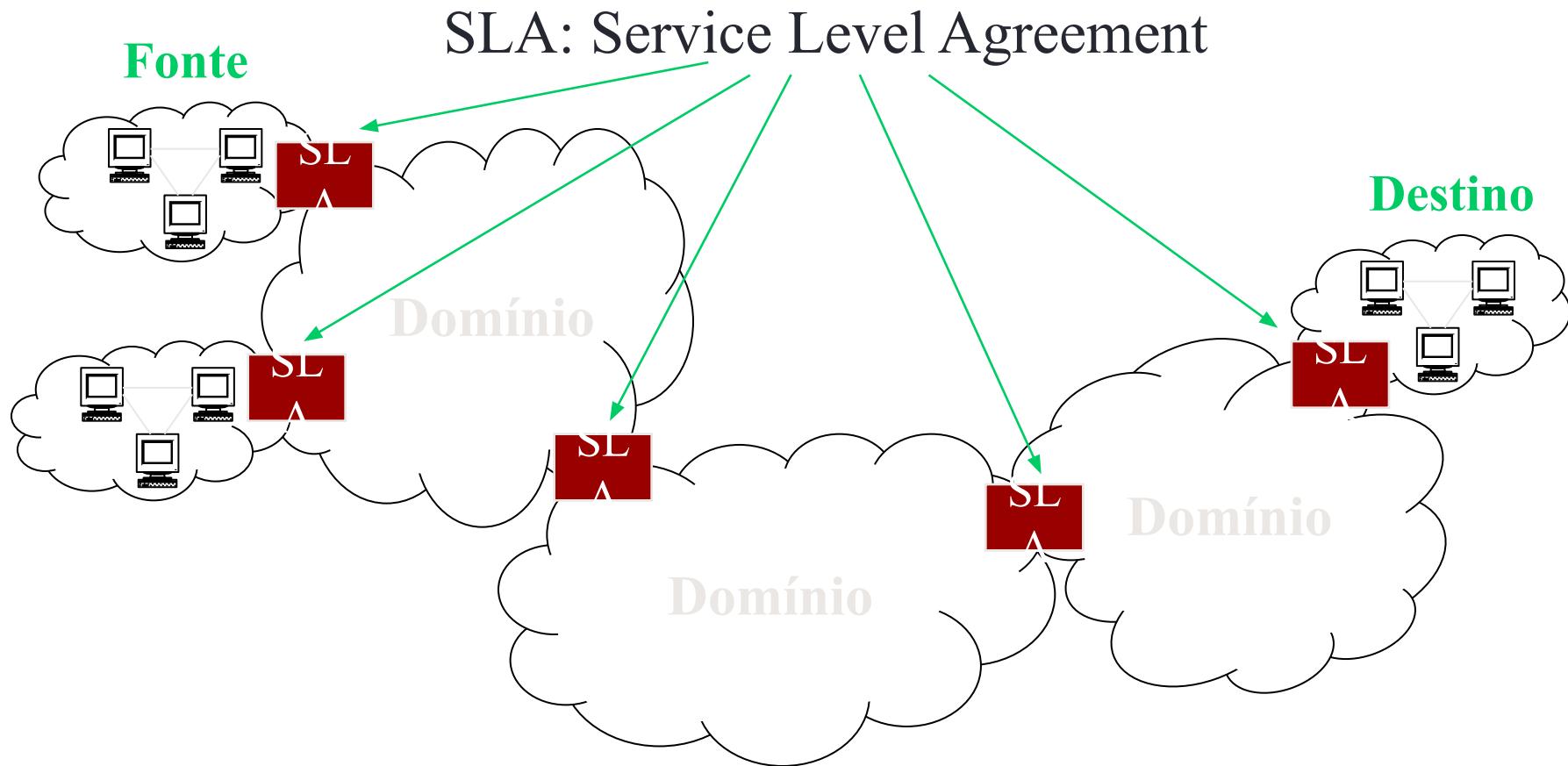
# Especificação do Nível de Serviço

## Parâmetros de Especificação de Serviços

- Âmbito: 1:1,1:N,...
- Identificação do Fluxo
- Descrição do tráfego
  - Características do tráfego e parâmetros de conformidade
  - Taxa de pico, parâmetros de token-bucket (b,r), MTU min/max
- Tratamento ao excesso
  - Que fazer ao tráfego fora de perfil: descarte, remarcação, calibração?
- Desempenho: débito, atraso, perda, jitter
  - Garantias qualitativas e quantitativas
- Temporização: activação/cancelamento do serviço
- Fiabilidade: tempo entre falhas

# Arquitetura de Serviços Diferenciados

## *Modelo Lógico*

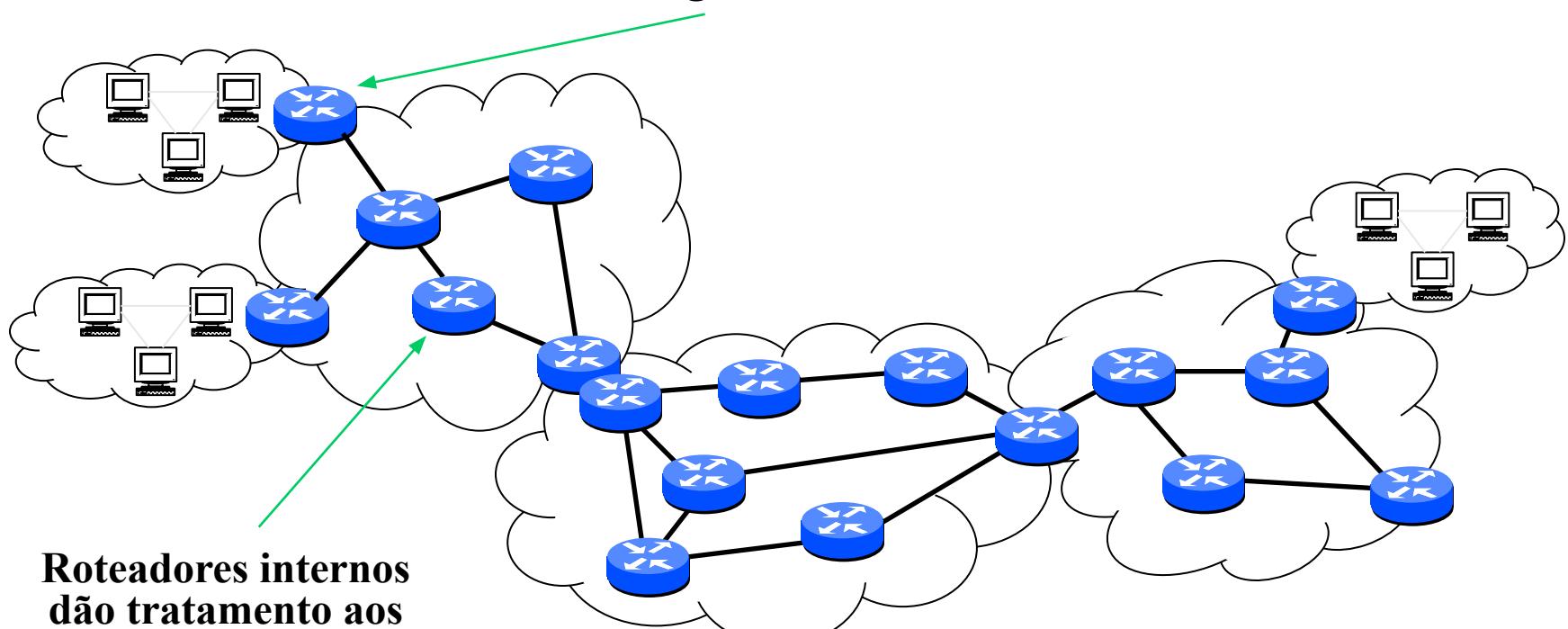


Domínios proporcionam serviços especificados no SLA aos seus clientes

# Arquitetura de Serviços Diferenciados

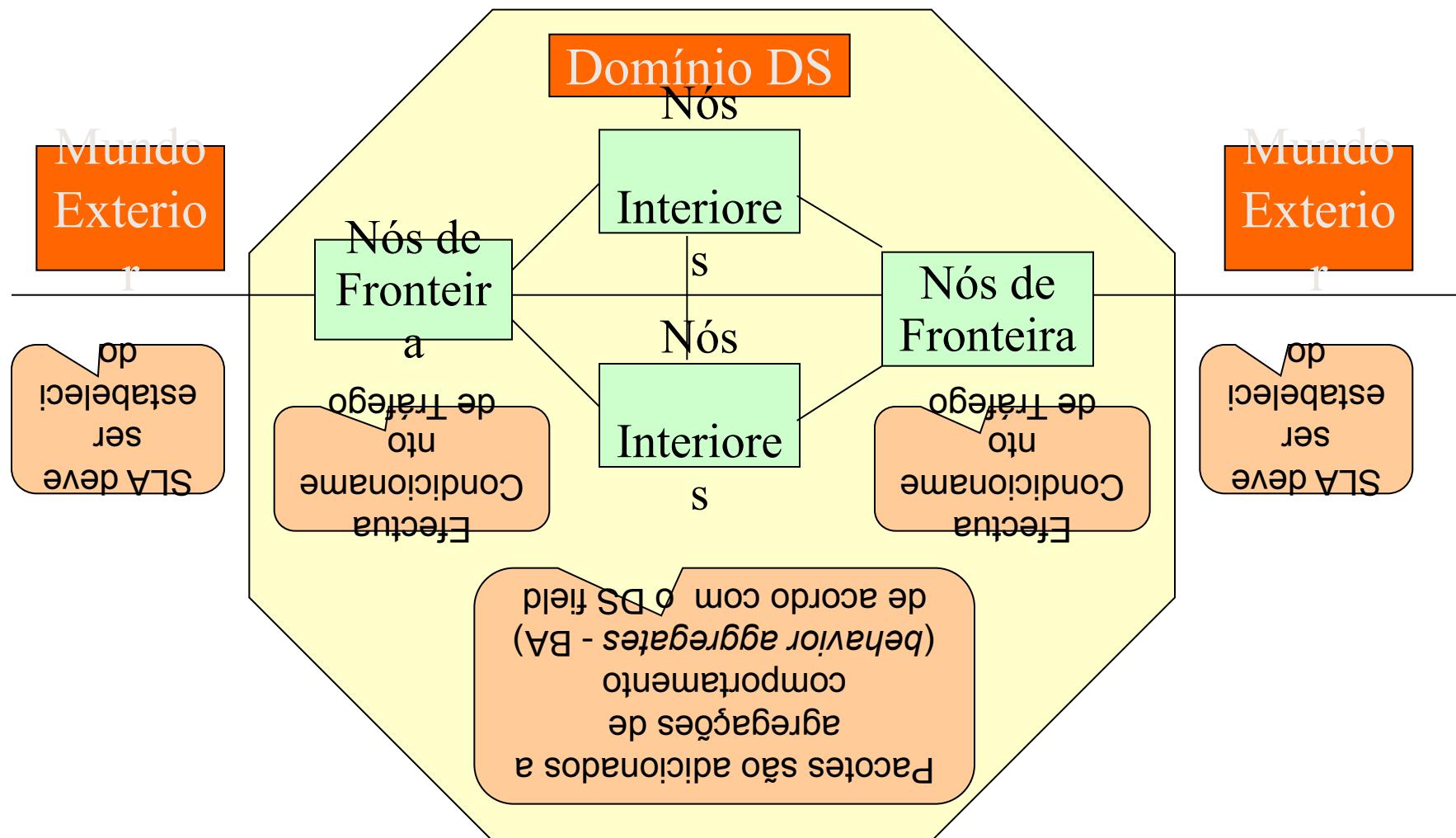
## *Modelo Físico*

**Política de tráfego nos roteadores de fronteira**



**Roteadores internos  
dão tratamento aos  
pacotes de acordo  
com o PHB indicado  
no DSCP**

# Arquitetura de Serviços Diferenciados



# Serviços Diferenciados



# Serviços Diferenciados

De acordo com as lições aprendidas?

- complexidade (estado por fluxo) na periferia da rede
  - leaky bucket marking
- Roteadores de núcleo de alta velocidade
  - 1-bit determina o comportamento na expedição
- **Sobre-alocação de largura de banda** para tráfego dentro do perfil e melhor esforço para tráfego fora do perfil

# Per-Hop Behavior (PHB)

- Tratamento de expedição que os pacotes recebem nos Roteadores
- Descreve o comportamento na expedição do tráfego agregado num nó DS
  - Pacotes com o mesmo DCSP
- Determina estratégia de alocação de recursos para construção de serviços diferenciados
- É especificado em termos de
  - Prioridade na obtenção de recursos (buffers, largura de banda) e características de tráfego (atraso, perdas)
- É concretizado por mecanismos de
  - Gestão de Filas (ex. RED)
  - Escalonamento (ex. WFQ, CBQ,...)

# PHBs Normalizados

- PHB **EF** (Expedited Forwarding) (RFC 2598)
  - Expedição expresso (acelerada)
  - Baixa perda, atraso e variação do atraso (*jitter*)
  - Preferência total de encaminhamento
- PHB **AF** (Assured Forwarding) (RFC 2597)
  - Grupo de PHBs de expedição assegurada
  - 4 classes de serviços com 3 níveis de descarte
  - Define tratamentos diferenciados aos pacotes, do tipo “melhor que o melhor esforço”

# PHB EF (Expedited Forwarding)

- Serviço com elevada fiabilidade
  - Baixa probabilidade de perda
  - Atraso e jitter reduzidos
  - Largura de banda garantida
- Condicionamento de tráfego da classe
  - Débito máximo de chegada < débito mínimo de partida
- Simula uma conexão ponto-a-ponto ou uma linha virtual fim-a-fim
- Limitação de interferência com outros tráfegos
  - Excedente de tráfego é rejeitado
  - Serviço de acesso limitado
  - Elevado custo

# PHB AF (Assured Forwarding)

- Garantia de expedição
  - Tráfego dentro do perfil tem grande probabilidade
  - Tráfego fora do perfil tem menor probabilidade
- Sem garantias de limite no atraso e jitter
- Nível de garantia de expedição dum pacote depende de
  - Recursos alocados à classe AF
  - Carga actual da classe
  - Prioridade de rejeição do pacote em situação de congestão

# PHB AF (Assured Forwarding)

- 4 classes AF independentes
- 4 níveis de alocação de recursos (buffers e largura de banda)
- 4 níveis de garantia de expedição
- 3 níveis de precedência de rejeição por classe, em situação de congestão
- Possibilidade de partilha de recursos entre diferentes classes AF

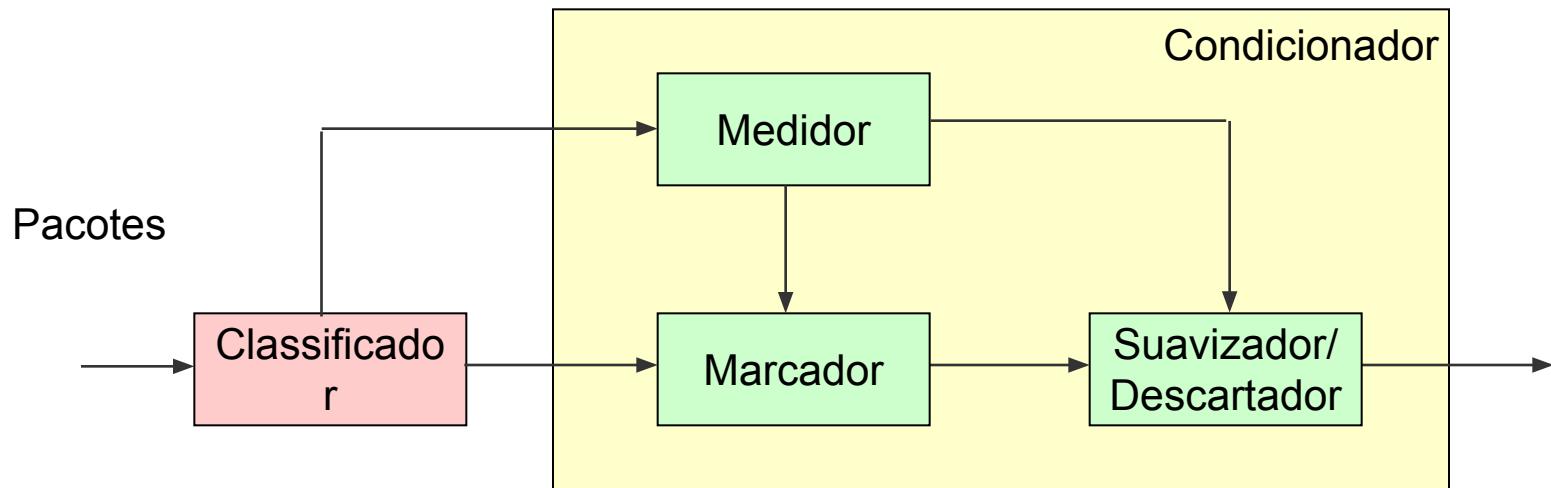
# Exemplos de Serviços

- Serviço Premium (PHB EF)
  - emulação de linha dedicada a uma taxa de pico especificada
- Serviço Assegurado (PHB AF)
  - a rede parece estar “levemente carregada” para tráfego em perfil especificado (taxa e rajada)
- Serviço Olímpico (PHB AF)
  - serviço “melhor” **relativo** a quem paga menos
  - semelhante ao serviço assegurado, mas com três classes de serviços: ouro, prata e bronze

# Condicionamento de Tráfego(1)

- **Classificador**: selecciona os pacotes dentro de um fluxo através do cabeçalho (BA e MF)
- **Medidor**: mede o fluxo para verificar se está de acordo com o perfil de tráfego contratado (SLA)
- **Marcador**: grava determinado padrão de bits no *codepoint (DSCP)*
- **Suavizador**: atrasa tráfego fora do perfil, para torná-lo dentro do perfil

# Condicionamento de Tráfego(2)



- ☐ Nem todos os quatro elementos precisam estar presentes em todos os nós de fronteira

# Política de Tráfego

## □ Objectivo

- Verificação ou condicionamento de tráfego
- Nos serviços diferenciados, corresponde a detectar o perfil de tráfego
  - Remarcação ou descarte de tráfego fora de perfil

## □ Mecanismos

- Sistema de créditos
- Medição contínua da taxa de transmissão

# Filas e Escalonamento(1)

## Gestão de Filas

- Prestar serviços diferenciados envolve
  - Colocar pacotes em filas diferentes para cumprir requisitos de QoS
- Mecanismos
  - FIFO
  - Filas com prioridade (Priority Queuing)
    - Servir primeiro tráfego com prioridade
    - Abordagem inicial para diferenciar tráfego
    - Sobrecarga computacional quando concretizado com fila única
    - Pode monopolizar recursos

# Filas e Escalonamento(2)

## Gestão de Filas

- Filas baseadas em classes
  - Class Based Queuing (CBQ)
  - i filas com prioridades distintas
  - $x_i$  bytes de cada fila numa rotação
  - Resolve o problema da negação de recursos do modelo anterior
- Filas ponderadas de acordo com a carga
  - Weighted Fair Queuing (WFQ)
  - Filas distintas com base na prioridade da fila e volume de tráfego
  - Tempos de resposta previsíveis a utilizadores pesados e menos pesados
  - Fluxos de baixo volume têm tratamento preferencial
  - Fluxos de alto volume partilham a largura de banda restante
  - Evita a *starvation* do buffer e da largura de banda

# Filas e Escalonamento (3)

## Gestão de Filas: descarte de pacotes

- Objectivos

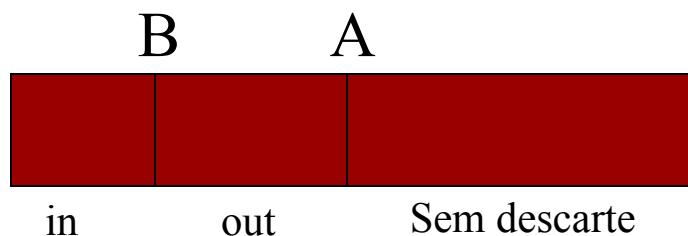
- Controlar o comprimento das filas de espera
- Evitar a congestão
  - Descartar pacotes
  - As fontes detectam perdas e abrandam a transmissão

- Mecanismos

- RED (Random Early Detection)
- RIO (Random Early Decetection with In and Out)

# Filas e Escalonamento (4)

- RED- Random Early Detection
  - Descarta pacotes aleatoriamente
- RIO – Random Early Dectection with In and Out (RIO)
  - Algoritmo RED distinto para pacotes *dentro do perfil* e *fora de perfil*



# Calibrador de Tráfego

## *Traffic Shaper*

### □ Objectivos

- Controlar o volume e taxa de saída de tráfego
- DiffServ: Regulação de tráfego agregado de saída nos nós de egresso

### □ Mecanismos utilizados

#### □ Leaky-bucket

- Fluxo regular de saída (configurável, em bits/seg)
- Elimina explosões (bursts) de tráfego
- Ineficiente utilização de recursos

#### □ Token-bucket

- Transmissão baseada na existência de tokens no bucket (enquanto houver tokens pode transmitir)
- Permite explosões de tráfego até um dado limite

# Controle de Admissão

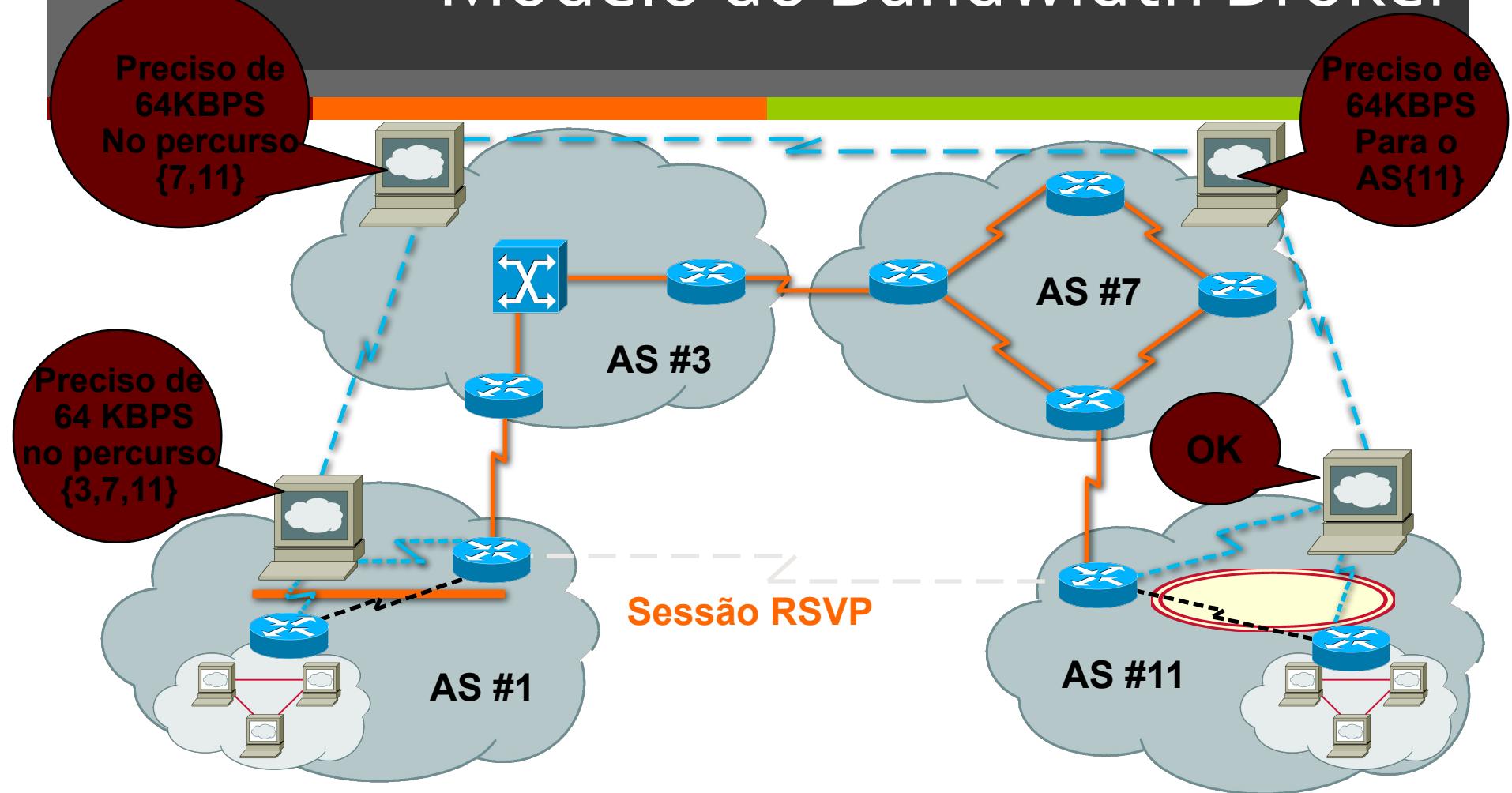
- Como gerenciar o nível de serviço contratado?
- Como garantir que os recursos disponíveis não são ultrapassados?
  - Controlar o tráfego que é admitido
  - Garantir que o novo fluxo não prejudica os existentes
    - Estratégias centralizadas e distribuídas
    - Acções no domínio fonte ou no ISP
    - Comunicação entre domínios
  - Compromisso entre complexidade e grau de garantia

# Controle de admissão

## Estratégia centralizada: Bandwidth Broker

- Funções Intra-Domínio
  - Gestão de Recursos do domínio
  - Controle de admissão de fluxos
    - Decisão local ou fim-a-fim
  - Configuração de Sistemas
    - Aplicar regras de classificação e condicionamento
- Funções Inter-Domínio
  - Comunicação com BBs doutros domínios
    - Para processo de admissão e reservas fim-a-fim
    - Protocolo de sinalização entre BBs de domínios diferentes

# Modelo do Bandwidth Broker



# Controle de admissão

## Estratégias distribuídas

- Controle de admissão nos extremos
  - Extremos: hosts ou Roteadores de fronteira
  - Decisão baseada em
    - Medição activa do QoS (probing)
    - Medição passiva do QoS (LB, atraso, perdas)
- Controle de admissão em cada nó
  - Em conformidade com estado da classe
- Compromisso entre a complexidade e grau de garantia

# Alocação de Serviços e Recursos

## Domínio Fonte

- Partilha dos recursos e serviços do SLA entre máquinas e aplicações
- Opções
  - Decisão dos hosts ou Roteadores de fronteira
    - Com ou sem probing
  - Decisão do BB
  - Melhor Esforço

# Alocação de Serviços e Recursos

## Domínio do Fornecedor de Serviço

- Configuração dos Roteadores de fronteira
  - Nível de serviço estático
    - Alocação estática de recursos a cada cliente
  - Nível de serviço dinâmico
    - Alocação baseada em sinalização
      - Ex: BB cliente comunica com BB do ISP
  - Sem nível de serviço
    - Best Effort provavelmente

# Comportamento por domínio

## *Per-Domain-Behaviour (PDB)*

- Tratamento da expedição dum agregado de tráfego no percurso dum domínio DS
- Influenciado pelos Roteadores de fronteira escolhidos
- Configuração de cada PHB
- Medição dos parâmetros permite
  - Estabelecer SLA na fronteira do domínio DS

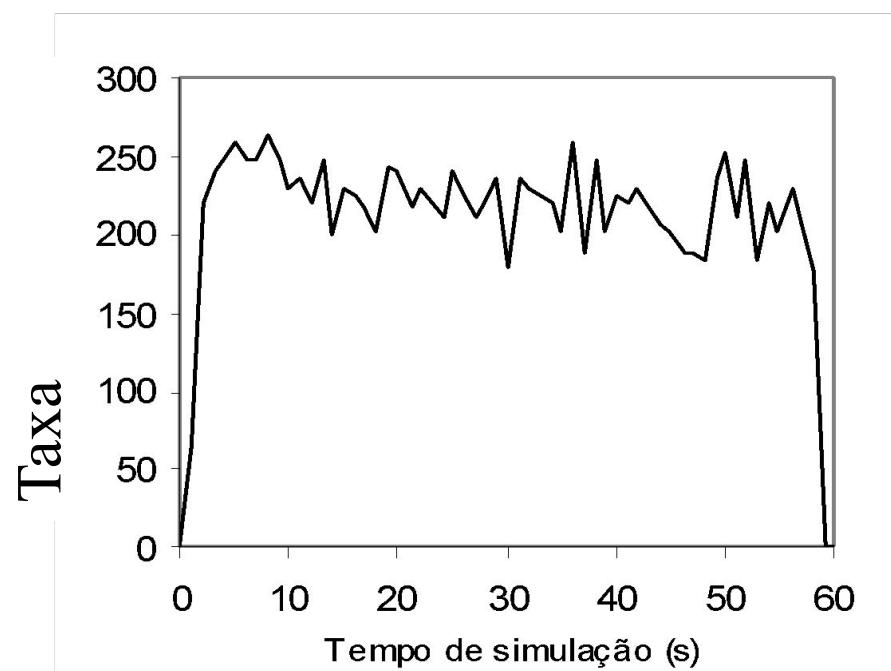
# Comportamento por domínio

## *Per-Domain-Behaviour (PDB)*

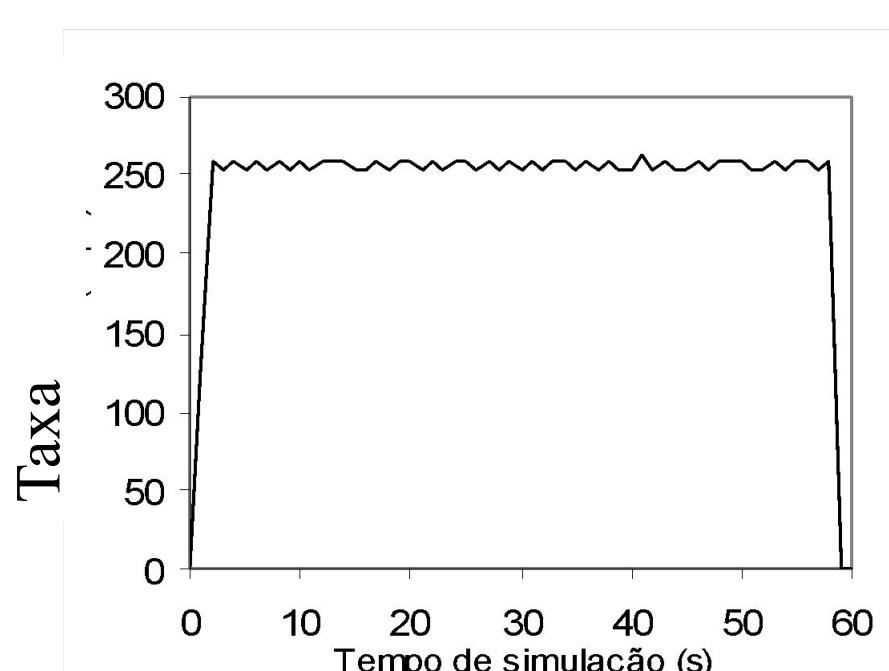
- Especificações do IETF
  - Virtual Wire PDB
    - Tráfego agregado tratado ao nível de linha dedicada
  - Assured Rate PDB
    - Tráfego agregado com pequena probabilidade de perda se a taxa for inferior a dado limite
  - Bulk Handling PDB
    - Tráfego agregado com baixa prioridade inferior ao melhor esforço

# Simulação DiffServ

A) Sem DiffServ

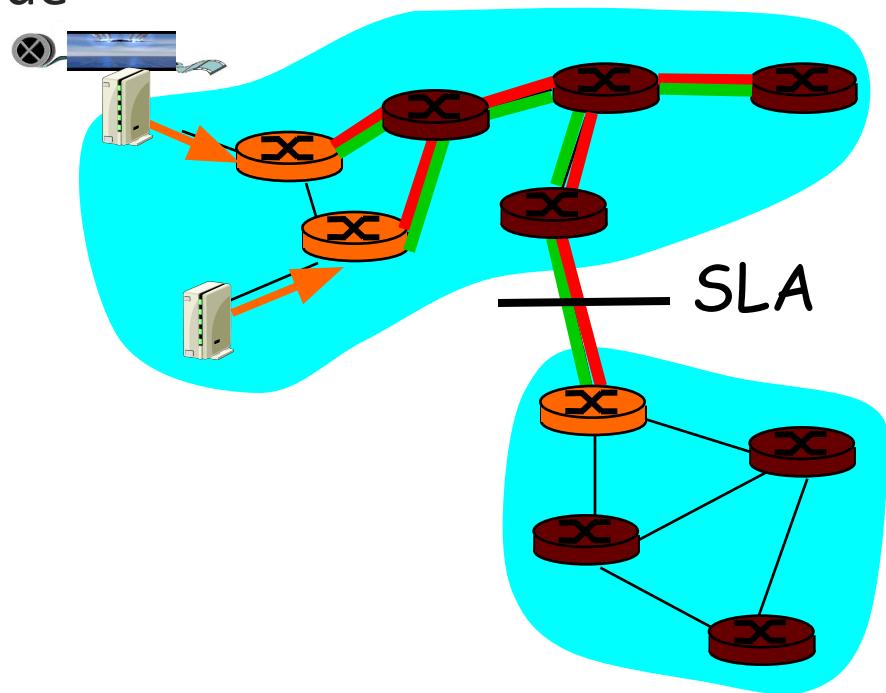


B) Com DiffServ



# Serviços Diferenciados: desafios

- Agregação e desagregação de fluxos marcados
- Aspectos fim-a-fim
  - SLA's inter-domínio
  - bandwidth brokers
- Monitorização, dimensionamento
  - Quanto deve ser o sobre-dimensionamento ?
- Protocolos de alto-nível, (i.e., TCP), desempenho da aplicação



# Diferenciados versus Integrados

| Parâmetro de comparação               | SD                | SI          |
|---------------------------------------|-------------------|-------------|
| Escalabilidade                        | Melhor            | Pior        |
| Protocolo de sinalização              | Sim               | Não         |
| Granularidade de alocação de recursos | Classe            | fluxo       |
| Informação de estado                  | O(classes)        | O(fluxos)   |
| Agregação de fluxos em classes        | Sim               | Prevista    |
| Acções de Controle de tráfego         | Periferia da rede | Toda a rede |
| Alteração na base instalada           | Mínima            | Grande      |
| Implementação global                  | Acessível         | Difícil     |
| Garantia de QoS                       | Pior              | Melhor      |

# Pergunta

- Compare sucintamente o modelo dos serviços integrados e diferenciados usados para fornecer serviços de rede com QoS usando os seguintes parâmetros: escalabilidade, protocolo de sinalização, granularidade na alocação de recursos, manutenção de informação de estado, alteração da base instalada, nível de garantia de QoS.

# Resposta (1)

- Os serviços integrados têm grande problemas de escala porque são orientados ao fluxo. O mesmo não acontece com os serviços diferenciados porque funcionam sim com fluxos agrupados em (poucas) classes. Nos serviços integrados a filosofia é a reserva por fluxo pelo que é necessário haver sinalização antes de cada ligação. No modelo de serviços diferenciados se houver algum dinamismo pode ser usada alguma sinalização mas de forma mais coordenada e não como condição prévia de estabelecimento de ligação ou de envio dos dados para a rede.

# Resposta(2)

- Podemos dizer que no caso dos serviços diferenciados a alocação de recursos é feita por classe enquanto nos integrados é feita por fluxo pelo que é necessário manter informação de estado por cada fluxo. No caso dos serviços diferenciados só há que manter contabilidade dos recursos usados por classe.
- Relativamente à base instalada, ao contrário dos serviços integrados os serviços diferenciados precisam de poucas modificações no núcleo da rede já que as funcionalidades mais complexas são colocadas na periferia.
- Os serviços integrados naturalmente oferecem um nível de garantia de QoS bastante superior.