Análise de Desempenho de Mecanismos de Diferenciação de Serviços para Tráfegos de Voz na Internet*

Artur Ziviani
artur@gta.ufrj.br

José F. de Rezende
rezende@gta.ufrj.br

Otto Carlos M. B. Duarte
otto@gta.ufrj.br

Grupo de Teleinformática e Automação – GTA
Programa de Engenharia Elétrica – COPPE
Universidade Federal do Rio de Janeiro
C. P. 68504 - 21945-970 - Rio de Janeiro - RJ - Brasil

Resumo

O objetivo da arquitetura de diferenciação de serviços é a provisão de QoS na Internet de forma escalável. Uma análise dos serviços prestados para o transporte de tráfegos de voz pelos esquemas de Encaminhamento Expresso e Encaminhamento Assegurado é realizada discutindo-se o atraso e o *jitter* observados. Os resultados demonstram ser possível obter um ganho estatístico de aproximadamente 70% preservando-se a qualidade nos fluxos de voz quando há o transporte combinado de pacotes de voz e de melhor esforço.

Abstract

The main goal of the differentiated services architecture is the scalable provision of QoS in the Internet. This paper evaluates the performance achieved by voice traffic when using the Expedited and Assured Forwarding services in terms of delay and jitter. The results demonstrate that when voice and best-effort traffic are mixed, a statistical gain in the order of 70% can be achieved, while preserving the QoS required by the voice flows.

*Este trabalho foi realizado com recursos da UFRJ, FUJB, CNPq, CAPES, COFECUB, FAPERJ e REENGE.

1 Introdução

Atualmente, a Internet apresenta-se como uma infraestrutura de base para a sociedade de informação globalizada, onde o número de estações conectadas apresenta um crescimento vertiginoso, com um conseqüente aumento de tráfego. No entanto, o modelo de melhor esforço adotado pelo protocolo IP para o envio de informações não atende as exigências de Qualidade de Serviço (*Quality of Service* - QoS) das diferentes mídias de informação. Este artigo aborda o problema do suporte aos requisitos de QoS da mídia de voz neste contexto.

Para fornecer diferentes serviços na Internet, duas abordagens distintas foram propostas nos últimos anos. A primeira, desenvolvida pelo grupo de trabalho *Integrated Services* do IETF (*Internet Engineering Task Force*), consiste na definição de uma arquitetura de integração de serviços, chamada arquitetura IIS (*Internet Integrated Services*) [2]. O ponto mais relevante desta arquitetura refere-se à reserva de recursos nos nós da rede para atender aos requisitos de QoS de um determinado fluxo. Desta forma, o estado de cada fluxo deve ser mantido pelos roteadores, o que significa uma grande quantidade de estados nos roteadores do "núcleo" (*core*) da rede que transportam um número muito elevado de fluxos. Estes estados e o processamento a eles associado acarretam uma sobrecarga destes roteadores, que passam a ser um "gargalo" no desempenho. A segunda abordagem, desenvolvida pelo grupo de trabalho *Differentiated Services* do IETF, busca prover a diferenciação de serviços na Internet sem a necessidade de um estado para cada fluxo e de sinalização a cada nó, aumentando assim a sua escalabilidade. Assim, esta solução pode ser utilizada nos roteadores do núcleo da rede. A idéia básica desta proposta é prover diferentes tratamentos às agregações de fluxos no interior de um domínio de diferenciação de serviços [9, 1].

O tratamento em cada nó da rede é denominado de Comportamento por Nó (*Per Hop Behavior* - PHB). Dois PHBs estão atualmente em ampla discussão dentro da proposta de Diferenciação de Serviços (DS), são eles, o PHB de Encaminhamento Expresso (*Expedited Forwarding* - EF) [7] e o PHB de Encaminhamento Assegurado (*Assured Forwarding* - AF) [6]. Em dois trabalhos anteriores [4, 10] foram avaliados individualmente cada um dos PHBs propostos, medindo o nível de serviço fornecido por cada um deles em presença de tráfegos de melhor esforço. Este artigo avalia o desempenho de fluxos de voz que atravessam um domínio DS que implementa tanto o PHB de Encaminhamento Expresso quanto o PHB de Encaminhamento Assegurado. O objetivo principal é avaliar a interação entre estes PHBs quando implementados numa mesma infraestrutura. Estes serviços foram implementados como uma extensão ao simulador ns-2 [8]. Para tanto foi necessária a implementação de mecanismos de escalonamento de pacotes, de condicionamento de tráfego, tais como marcadores de pacotes, suavizadores e policiadores de tráfego; e ainda, mecanismos de gerenciamento ativo de filas para a implementação do PHB AF. Os resultados mostram que com uma provisão adequada de recursos e com uma parametrização apropriada dos mecanismos envolvidos, ambos PHBs são capazes de fornecer o nível de serviço exigido pelos fluxos de voz. Um ganho estatístico da ordem de 70% foi observado no modelo de simulação avaliado. Este ganho em banda passante permite absorver um maior número de tráfegos de voz, além de atender o tráfego de melhor esforço.

O artigo está estruturado da seguinte forma. Na seção 2 são revistos os conceitos básicos da proposta de diferenciação de serviços. São também discutidas as características dos PHBs de Encaminhamento Expresso e Encaminhamento Assegurado. A seção 3 apresenta os modelos adotados nas simulações realizadas e analisa os resultados obtidos. Finalmente, na seção 4, são apresentadas as conclusões e as perspectivas deste trabalho.

2 Serviços Diferenciados

A proposta de diferenciação de serviços baseia-se na utilização do campo ToS (*Type of Service*) do cabeçalho IP como indicador do tratamento a ser recebido por uma agregação de fluxos no interior de um domínio DS. A classificação de fluxos individuais em agregações, a verificação de conformidade ao perfil de tráfego contratado e a marcação dos pacotes são atividades complexas que devem ser realizadas no entorno do domínio DS. Assim, no interior do domínio DS, a simples análise do campo DS indica qual PHB deve tratar o pacote no seu encaminhamento. Desta forma, atividades complexas são mantidas nas fronteiras do domínio DS, dispensando a necessidade de manutenção de um estado por fluxo no interior do domínio. O serviço fim-a-fim obtido num domínio DS é o resultado da combinação dos PHBs recebidos ao longo da trajetória do fluxo pelo interior do domínio.

2.1 Encaminhamento Expresso

O modelo adotado pelo PHB EF visa prover um serviço com baixo retardo e com pequenas variações neste retardo (*jitter*) [7]. Uma determinada banda passante da rede é reservada para o atendimento de uma agregação de fluxos cuja taxa de serviço em cada nó é dada por um valor mínimo bem definido, independente da intensidade dos outros tráfegos transitando neste nó. Os fluxos que compõem esse agregado devem ser limitados por uma taxa de pico. A agregação de fluxos ocorre na fronteira de um domínio DS que implementa este modelo. Um mecanismo de policiamento atua neste ponto, marcando os pacotes que receberão o tratamento EF no interior do domínio DS de acordo com o taxa de pico contratada por cada usuário. Os roteadores internos do domínio policiam a agregação de fluxos, descartando os pacotes que não estiverem de acordo com as características do tráfego agregado. O inconveniente desta proposta é justamente alocar os recursos de uma forma radical, onde a capacidade da rede como um todo é drasticamente reduzida pelo que foi alocado aos fluxos que utilizam o serviço EF. Por conseqüência, todos os problemas de congestionamento somente afetarão os tráfegos de melhor esforço.

A implementação do PHB EF pode ser realizada por diferentes mecanismos de escalonamento de filas. Um mecanismo de fila prioritária (*Priority Round Robin* - PRR) exerce o comportamento desejado. O policiamento nas fronteiras do domínio DS torna-se especialmente importante neste caso, pois um tráfego abusivo pode obter recursos não contratados, em detrimento dos demais fluxos. Uma alternativa é a adoção de um mecanismo com filas atendidas de forma ponderada, tal como o *weighted round robin* (WRR), onde a fatia de serviço alocada à fila EF deve ser superior a taxa de entrada da agregação EF. Jacobson *et al.* [7] analisam como implementações do serviço EF por PRR ou por WRR comportam-se em termos da variação no retardo (*jitter*). Os resultados demonstram que a implementação por PRR é sempre o melhor caso, pois os pacotes do tráfego EF somente são atrasados pela presença de outros pacotes EF em sua fila. No entanto, os roteadores comerciais atuais já implementam o mecanismo de WRR, o que justifica a investigação de alternativas para a melhoria de seu desempenho.

2.2 Encaminhamento Assegurado

O PHB AF é definido em [6]. Este modelo de serviço, ao invés de fornecer uma garantia estrita, fornece uma expectativa de serviço que será obtida por um determinado tráfego quando existirem momentos de congestionamento. O controle de admissão exerce papel importante neste modelo, pois ele é encarregado de garantir a existência dos recursos necessários (*buffers* e banda passante) durante todo o percurso da origem ao destino. Um perfil associado a cada tráfego define o serviço esperado por este. Um mecanismo de policiamento atua nas fronteiras do domínio DS de forma a marcar os pacotes em conformidade com este perfil. O descarte será prioritário para pacotes marcados como não-conformes, enquanto que o tráfego que esteja em conformidade recebe níveis de serviço previsíveis. Conforme os pacotes são encaminhados pela rede e agregados a outros fluxos os roteadores na fronteira de outras regiões somente policiam esta agregação de fluxos. Uma implementação do PHB AF visa minimizar os congestionamentos duradouros, porém permite congestionamentos curtos resultantes de rajadas. Isto é, o mecanismo de gerenciamento de filas deve detectar e reagir a congestionamentos duradouros dentro de cada classe e tolerar rajadas de pacotes que provocam congestionamentos curtos. Tal comportamento pode ser obtido por um mecanismo de gerenciamento ativo de filas como o RED (*Random Early Detection*) [5].

O RED é implementado nos roteadores e projetado para atuar em conjunto com o protocolo TCP. Através do RED é realizada a monitoração do tamanho da fila no roteador e caso hajam sinais de um congestionamento iminente, as fontes são notificadas de forma implícita pelo descarte de um de seus pacotes. Quando um roteador RED utiliza uma fila FIFO, ao invés de esperar que a fila se torne completamente cheia, o que implica no descarte de todos os pacotes que chegam, ele decide descartar cada novo pacote com uma certa probabilidade sempre que o tamanho da fila exceda um determinado nível. O roteador computa o tamanho médio da fila ($avglen$) utilizando uma média ponderada por uma variável q_{weight} de maneira similar àquela utilizada na computação dos temporizadores do TCP. A variável q_{weight} determina a tolerância do roteador RED às rajadas de longa duração. Ou seja, quanto maior for o q_{weight}, maior será o peso do tamanho instantâneo da fila no cálculo do novo valor do $avglen$. O RED utiliza dois patamares para determinar a probabilidade de descarte do próximo pacote a chegar na fila: *min_thresh* e *max_thresh*. Enquanto o tamanho médio da fila ($avglen$) permanecer abaixo de *min_thresh*, todos os pacotes que chegam ao roteador são aceitos na fila. Quando $avglen$ está entre os dois patamares, há uma probabilidade p do novo pacote ser descartado, onde p é uma função de $avglen$ e do tempo decorrido desde o descarte do último pacote. Se $avglen$ ultrapassa o limite *max_tresh*, todos os próximos pacotes serão descartados até a média de ocupação da fila se reduzir. A probabilidade de descarte tem seu limite superior determinado pelo parâmetro RED P_{max}.

O gerenciamento de *buffers* nos roteadores intermediários que implementam o serviço de Encaminhamento Assegurado é normalmente realizado pela adoção de dois algoritmos RED: um para pacotes que estão em conformidade com o perfil de tráfego definido e outro para os pacotes não-conformes. Este mecanismo é conhecido por RIO (RED com *IN* e OUT) [3]. O segundo mecanismo RED é configurado de forma mais agressiva que o primeiro, visando o descarte prioritário de pacotes OUT. O mecanismo RIO contabiliza o tamanho médio da fila para pacotes *IN* (avg_{IN}) e também o tamanho médio da fila considerando-se tanto pacotes *IN* como OUT (avg_{TOTAL}). A probabilidade de descarte de um pacote *IN* é função de avg_{IN}, enquanto a probabilidade de descarte de um pacote OUT é função de avg_{TOTAL}. A escolha adequada dos parâmetros (min_{IN}, max_{OUT}, P_{maxIN}) e (min_{OUT}, max_{OUT}, P_{maxOUT})

provê uma discriminação contra os pacotes *OUT*. Essencialmente, há o descarte inicial de pacotes *OUT* quando um congestionamento iminente é detectado e, caso o congestionamento persista, todos os pacotes *OUT* são descartados. Os pacotes *IN* somente são descartados quando há uma inundação de pacotes *IN*, o que numa rede bem dimensionada não deve ocorrer.

3 Resultados de Simulação

Este artigo tem como objetivo avaliar o desempenho de fluxos de voz atravessando redes que implementam a diferenciação de serviços através dos PHBs EF e AF. Para a manutenção de interatividade em uma conversação, assume-se que o retardo fim-a-fim do tráfego de voz deve estar abaixo de 250 ms. Uma baixa variação neste retardo (*jitter*) também é importante, pois diminui a quantidade de *buffers* nos receptores, necessários para compensar estas variações. Para medir o desempenho obtido por fluxos de voz, cada um desses PHBs foi implementado no simulador ns-2 [8]. Dentre os mecanismos implementados estão: mecanismos de escalonamento de pacotes (filas com prioridade, *weighted round-robin*, *weighted fair-queueing*), mecanismos de condicionamento de tráfego (suavizadores de tráfego e policiadores), mecanismos de marcação de pacotes de acordo com perfis de tráfego (*profilers*), e ainda mecanismos de gerenciamento ativo de filas (RIO [3]). Assim, pode-se implementar um PHB através da combinação de uma diversidade de mecanismos de marcação e escalonamento de pacotes e de condicionamento de tráfego.

Uma avaliação de cada um destes PHBs foi realizada tendo como objetivo determinar a adequação de cada um deles na provisão de garantias de QoS, em termos de banda passante, retardo e variação no retardo [4, 10]. Neste trabalho é avaliada a interação dos dois PHBs quando implementados simultaneamente num mesmo domínio DS.

3.1 Modelo de Simulação

A topologia de rede adotada nas simulações está representada na figura 1. Os nós numerados de 0 a 7 constituem um domínio de diferenciação de serviços e cada um deles é capaz de prover o PHB EF ou o PHB AF a uma determinada agregação de fluxos. Todos estes nós estão interligados por enlaces de 4,5 Mbps, com exceção do enlace de gargalo representado por um enlace de 384 kbps de longa distância. Os nós de 8 a 15 são as fontes de tráfego que devem ser tratadas diferenciadamente no interior do domínio DS. O tráfego gerado pelas fontes 8 e 9 compõem a agregação de fluxos EF. As fontes de 10 a 15 constituem a agregação de fluxos AF. Os nós de 16 a 34 contêm fontes de tráfego TCP, que formam o tráfego atendido pelo modelo de melhor esforço.

As fontes de 8 a 34 enviam seus pacotes para os destinos nos nós numerados de 35 a 61, respectivamente. As medidas realizadas referem-se a um fluxo de referência EF, cujos pacotes são gerados no nó 8 e destinam-se ao nó 35, e um fluxo de referência AF, cujos pacotes são gerados no nó 10 e destinam-se ao nó 37.

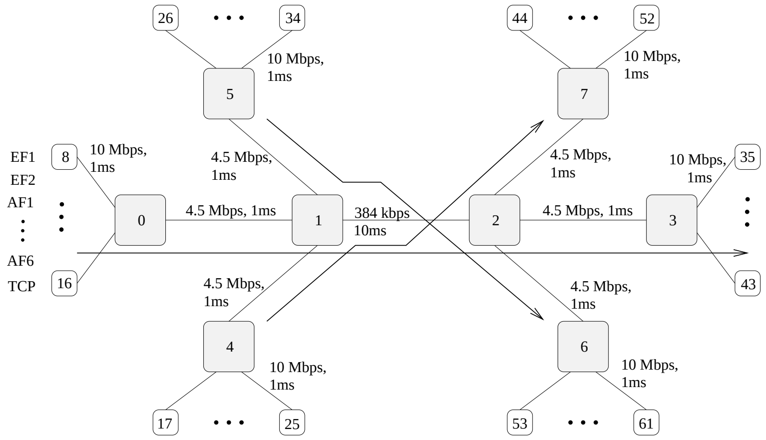


Figura 1: Topologia de simulação.

As fontes, tanto EF quanto AF, geram tráfego de voz seguindo o modelo de atividade-silêncio (On-Off). O intervalo de atividade (On) de cada fonte é determinado por uma variável aleatória exponencialmente distribuída de média 400 ms, onde é gerado tráfego de voz a uma taxa de 64 kbps. Os períodos de silêncio (Off) também são exponencialmente distribuídos de acordo com uma variável aleatória de média 600 ms. Logo, a taxa média de transmissão destas fontes é de 25,6 kbps. Os pacotes gerados pelas fontes On-Off possuem 576 octetos. As fontes TCP utilizam a implementação do TCP-Reno que adiciona o mecanismo de recuperação rápida aos mecanismos de controle de congestionamento originais do TCP. O tráfego TCP é gerado por aplicações FTP ativas durante toda a simulação.

Cada nó do domínio DS contém duas filas (EF e AF-BE) atendidas por um escalonador de pacotes WRR ou PRR. A primeira é uma fila FIFO que recebe a agregação de tráfego EF. A outra é uma fila com gerenciamento ativo do tipo RIO ocupada pela agregação AF (tanto pacotes *IN* quanto *OUT*) e pelo tráfego de melhor esforço (BE). Todos os pacotes do tráfego de melhor esforço são marcados como *OUT*. No caso de um escalonamento do tipo PRR, a fila EF é tratada prioritariamente em relação à fila AF-BE. Ou seja, somente são enviados pacotes da fila AF-BE quando a fila EF se encontra vazia. No caso do escalonamento WRR, uma porção relativa da banda passante é alocada a cada uma das filas. A estrutura interna de um nó do domínio DS é mostrada na figura 2.

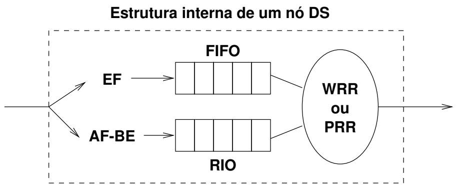


Figura 2: Estrutura interna de um nó DS.

3.2 Suavizadores

A agregação EF é constituída de dois fluxos oriundos dos nós 8 e 9. Como a taxa média de cada fluxo é de 25,6 kbps, a agregação possui 51,2 kbps de taxa média. Para a utilização da política de escalonamento de pacotes WRR, foi visto em [7] que se houver um equilíbrio exato entre a fatia de serviço WRR e o necessário à agregação EF no domínio DS, os resultados tornam-se instáveis. Jacobson *et al.* sugere a adoção de um fator de 1,06 entre a quantidade de banda passante solicitada pela agregação EF e a efetivamente reservada.

No caso da utilização do mecanismo de escalonamento WRR, um condicionamento do tráfego agregado EF na entrada do domínio DS pode ser necessário de acordo com a fatia de serviço a ser alocada a esta agregação. Para que a fatia reservada seja inferior à taxa de pico da agregação EF, esta agregação deve ser suavizada. A suavização permite por um lado controlar ou até mesmo eliminar as rajadas do tráfego agregado. Quando um mecanismo de balde furado (*leaky-bucket*) é utilizado com esta finalidade, a profundidade do balde determina o tamanho máximo de rajada permitido. Utilizando-se uma profundidade equivalente a um pacote, as rajadas são totalmente eliminadas. Neste caso, a fatia de serviço do WRR pode ser igual à taxa de reposição de créditos, ou seja inferior à taxa de pico. Quando um grande número de fluxos de voz compartilham os recursos, há a possibilidade de uma rajada acumulada sobrecarregar o gargalo, penalizando os demais tráfegos agregados. Além disso, a alocação pelo pico não permite o ganho estatístico no uso de fontes com taxa variável. No entanto, a suavização incorre num atraso adicional dos pacotes retidos no *buffer* de suavização. No caso da profundidade do balde ser maior do que o equivalente ao tamanho de um pacote, a fatia de serviço do WRR deve ser igual à soma das taxas de pico individuais. Sendo o mecanismo de escalonamento WRR do tipo conservativo, nos momentos de inatividade do tráfego EF as demais agregações utilizam a banda passante alocada a este tráfego. Com a limitação da rajada exercida pela profundidade do balde, limita-se o tempo máximo que a agregação EF utiliza a fatia de serviço a ela alocada.

Quando se utiliza o mecanismo de escalonamento PRR, a presença de rajadas não causa a violação da definição do serviço EF, desde que a taxa de pico da agregação seja inferior a capacidade do enlace de saída. Como discutido anteriormente, os recursos para o atendimento destas rajadas são utilizados em detrimento das demais agregações. A utilização de um suavizador para controlar o tamanho destas rajadas permite limitar o atraso sofrido pelas demais agregações.

Com o objetivo de estudar o atraso introduzido por um suavizador de tráfego que atue sobre uma agregação de fluxos, uma primeira simulação foi realizada. A suavização do tráfego agregado EF é implementada por um mecanismo de balde furado (*leaky bucket)*, associado a um *buffer*, localizado na entrada do domínio DS. Nesta simulação, as rajadas do tráfego de voz são totalmente suprimidas configurando-se a profundidade do balde em 4608 bits (1 pacote). A figura 3 apresenta o atraso médio e o percentil 95 do atraso introduzido pelo suavizador em função da taxa de reposição de créditos no balde. Os resultados mostram que a partir de uma taxa de reposição superior a 100 kbps no suavizador agregado, o atraso de 95% dos pacotes é inferior ao limite de 250 ms.

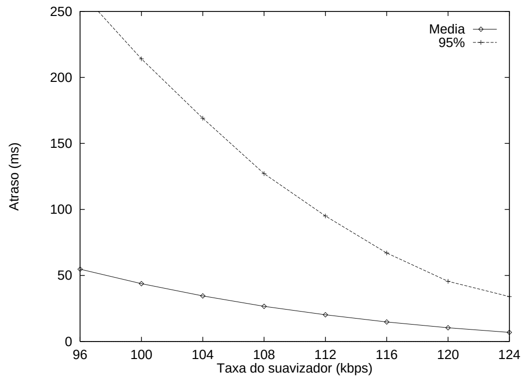


Figura 3: Variação da taxa de reposição de créditos no suavizador de tráfego agregado.

Numa segunda simulação, é utilizado um suavizador com a profundidade do balde igual a soma do tamanho médio da rajada de cada uma das fontes de tráfego EF. O tamanho médio da rajada de cada fonte EF equivale ao número de bits transmitidos no intervalo de atividade (25600 bits). Quando uma taxa superior à 80 kbps é utilizada, o atraso introduzido pelo suavizador é praticamente igual a zero.

3.3 Marcadores

Os pacotes gerados pelas seis fontes AF são marcados de acordo com um perfil de serviço contratado. Esta tarefa é realizada por marcadores (*profilers*) presentes nas próprias fontes ou na entrada do domínio DS. No primeiro caso, existe a necessidade de um policiador da agregação AF responsável pela remarcação de pacotes quando a taxa agregada AF excede a soma dos perfis individuais. Neste trabalho, estes marcadores são implementados também por um mecanismo de balde furado sem retenção de pacotes, onde os pacotes não conformes são marcados como OUT e imediatamente encaminhados. Estes marcadores estão localizados nas fontes e na entrada do domínio DS está localizado um policiador que atua sobre a agregação de fluxos AF. Os pacotes da agregação que não estiverem em conformidade com o perfil contratado pela agregação AF são remarcados como OUT e entram no domínio DS.

Os marcadores são configurados de forma que 90% do tráfego gerado pelas fontes EF sejam marcados como conformes. Os demais pacotes são marcados como OUT. O policiador na entrada do domínio DS é responsável pela remarcação de outros 2% do tráfego total. Para a obtenção da capacidade do balde do marcador que produz este resultado, a seguinte simulação foi realizada. A taxa de reposição de créditos no balde foi mantida constante em 25,6 kbps, enquanto a capacidade do balde era variada entre 1 e 40 pacotes. A figura 4 ilustra estes resultados. Foi verificado que com uma capacidade equivalente a 52200 bits, 90% dos pacotes são marcados com IN. Estes resultados preliminares são utilizados na configuração dos marcadores de pacotes AF e do policiador empregados em todas as simulações posteriores. O policiador tem sua taxa de reposição de créditos fixada em 153,6 kbps e a capacidade de seu balde estabelecida em 313200 bits. Estes valores correspondem a 6 vezes a configuração individual de cada marcador associado a uma fonte AF.

3.4 Tráfego de Voz

Nesta seção é avaliado o desempenho de fluxos de voz transportados pelos PHBs EF e AF, quando atuando em conjunto, através de um domínio DS. O domínio DS deve garantir aos fluxos de voz um atraso fim-a-fim máximo de 250 ms, atendendo também ao tráfego de melhor esforço, ainda que penalizado. O gerenciamento ativo da fila AF-BE é realizado pelo algoritmo RIO. Nas simulações realizadas os roteadores do domínio DS implementam o

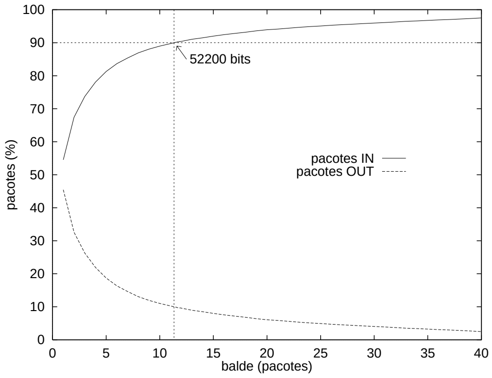


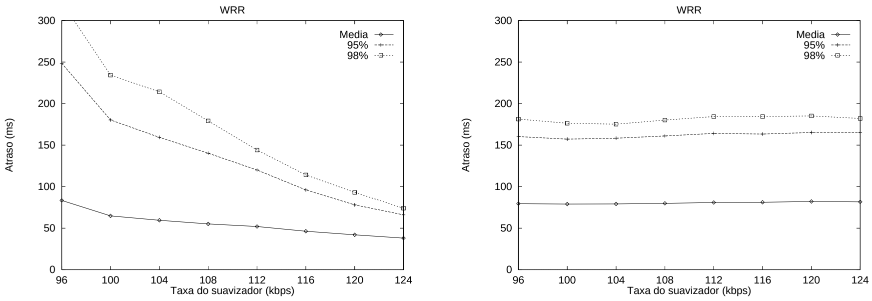
Figura 4: Marcação de pacotes em função da capacidade do balde.

algoritmo RIO com os parâmetros [4, 8, 0.002, 0.02][1] para os pacotes *IN* e [2, 4, 0.004, 0.1] para os pacotes *IN+OUT*.

A variação da taxa de reposição de créditos do suavizador de tráfego da agregação EF implica na alteração das fatias de serviço alocadas a cada agregação no escalonador WRR do interior do domínio DS. Quanto maior esta taxa, maior deve ser a porção relativa de serviço do WRR para atendê-la. Logo, há a diminuição da fatia alocada para o atendimento dos demais tráfegos, AF e melhor esforço. O objetivo da próxima simulação é avaliar o atraso dos fluxos de voz dentro do domínio DS, tanto utilizando o serviço EF quanto o serviço AF. Este atraso inclui o atraso introduzido pelo suavizador da agregação EF localizado na fronteira do domínio DS. As figuras 5(a) e 5(b) apresentam o atraso médio e o percentil 95 e 98 do atraso dos fluxos EF e AF de referência em função da taxa de reposição de créditos no suavizador para um escalonamento WRR, respectivamente.

O aumento da taxa de reposição de créditos no suavizador agregado diminui o atraso fim-a-fim observado na implementação por WRR, pois os pacotes EF permanecem menos tempo retidos no *buffer* do suavizador. No entanto, este aumento leva a uma maior fatia de serviço WRR alocada para a agregação EF, o que diminui a alocação para as demais agregações. Nos gráficos das figuras 5(a) e 5(b), a taxa de reposição de créditos do suavizador é variada a partir do valor para o qual o atraso do suavizador é inferior ao requisito de 250 ms para o atraso fim-a-fim, como visto na figura 3. Esta variação na taxa de reposição de créditos do suavizador penalizou em somente 4,3% o percentil

[1] $[min\_thresh,\ max\_thresh,\ q_{weight},\ P_{max}]$



(a) EF (b) AF

Figura 5: Atraso dos tráfegos EF e AF no WRR.

95 do atraso do tráfego AF. Esta pequena variação se deve ao fato de que o número de fluxos EF não é suficiente para infringir uma maior penalização. Quanto maior for o número de fluxos EF, espera-se uma maior degradação do serviço oferecido ao tráfego AF.

Já na implementação por PRR, como os pacotes são encaminhados de forma prioritária no interior do domínio DS, os atrasos observados por estes pacotes são decorrentes somente da presença de outros pacotes EF na fila prioritária. O fluxo de voz atendido pelo serviço AF sofre um atraso maior na implementação por PRR, o que pode ser creditado às rajadas permitidas ao fluxo EF atendido prioritariamente. Porém este acréscimo no atraso não compromete a qualidade ofertada pelo PHB AF ao fluxo de voz. Os resultados correspondentes aos fluxo de referência EF e AF são mostrados na tabela 1. Também foi medido o *goodput* total obtido pelo tráfego de melhor esforço. O desempenho dos fluxos TCP atendidos pelo serviço de melhor esforço tanto na implementação por WRR quanto na implementação por PRR é da ordem de 160,0 kbps.

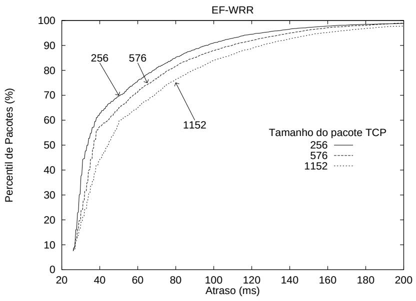
		Percentil de atraso (ms)	
	Atraso Médio (ms)	**95**	**98**
EF	32,29	38,12	40,00
AF	93,00	182,24	202,36

Tabela 1: Atraso do tráfego AF e EF no PRR.

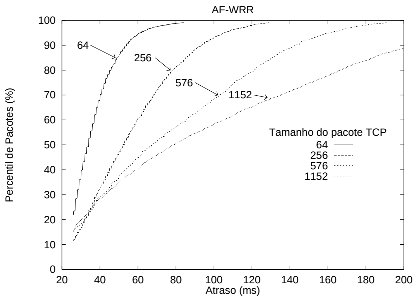
O enlace de gargalo do domínio DS pode suportar 6 chamadas telefônicas normais (64 kbps). Porém, com a arquitetura de diferenciação de serviços são suportadas 2 chamadas pelo serviço EF e 6 chamadas pelo serviço AF, num total de 512 kbps. Além dos tráfegos de voz, também são atendidos 19 fluxos TCP servidos pelo modelo de melhor esforço que podem alcançar um *goodput* total de 160 kbps, totalizando 672 kbps. Estes resultados mostram que um ganho estatístico em banda passante da ordem de 70% é obtido em relação ao enlace de gargalo de 384 kbps. Pode-se verificar, que configurando-se adequadamente os recursos de cada serviço, pelo menos 98% dos pacotes de voz atendidos pelos dois serviços sofrem atrasos inferiores a 250 ms, que é o limite de atraso fim-a-fim para que a interatividade de uma conversação seja preservada.

Na simulação seguinte é investigada a influência do tamanho do pacote do fluxo de melhor esforço no atraso dos fluxos de voz atendidos pelos serviços de EF e AF, quando escalonados por WRR. A figura 6(a) mostra o percentil de atraso dos pacotes pertencentes à agregação EF para diferentes tamanhos de pacote do tráfego de melhor esforço. O resultado correspondente para o pacotes da agregação AF é apresentado na figura 6(b). O atraso dos fluxos AF é dependente do tamanho do pacote do tráfego de melhor esforço que compartilha a fila AF-BE. A probabilidade de descarte do algoritmo RIO é aplicada no momento de chegada de um pacote. Quando um pacote é aceito na fila, seja *IN* ou *OUT*, este não pode ser mais descartado e não há o reordenamento de pacotes. Assim os pacotes presentes na fila aguardam o encaminhamento de todos os pacotes à sua frente para serem atendidos. Quanto maior for o pacote do fluxo de melhor esforço, maior é o seu tempo de transmissão, o que acarreta um maior atraso para pacotes de voz AF presentes na fila. Pode-se notar que o atraso do tráfego EF não é significativamente influenciado pelo tamanho do pacote do tráfego de melhor esforço, pois a agregação EF possui uma fatia exclusiva de atendimento pelo escalonador WRR.

A influência do tempo de transmissão dos pacotes presentes na fila AF-BE também reflete-se no *jitter*. Esta influência é observada em função do número de pacotes presentes na fila AF-BE quando um novo pacote é colocado na fila. A figura 7 apresenta o *jitter* observado por um fluxo de voz AF atendido por um escalonador WRR. A taxa de reposição de créditos requisitada ao suavizador agregado é de 96 kbps. O tamanho do pacote TCP é de 576 octetos, o mesmo dos fluxos de voz AF. Pode-se verificar que grupos de pacotes possuem o mesmo *jitter*, o que é ilustrado pelos patamares presentes na figura 7. Os valores de *jitter* apresentados por estes grupos de pacotes estão espaçados por intervalos de 12 ms, equivalente ao tempo de transmissão de um pacote de 576 octetos a uma taxa de 384 kbps.



(a) EF



(b) AF

Figura 6: Atraso dos tráfegos EF e AF em função do tamanho do pacote TCP.

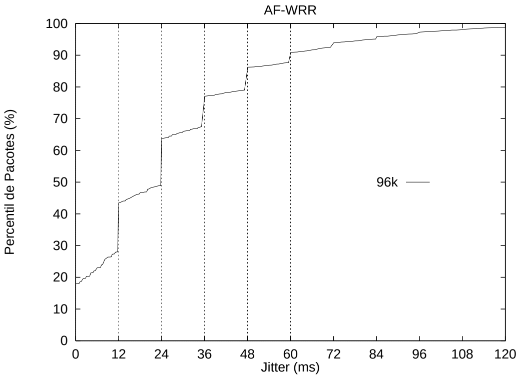


Figura 7: *Jitter* do tráfego AF.

4 Conclusão

Neste artigo foi apresentado um estudo de simulação sobre o tréfego de voz atendido por uma arquitetura de diferenciação de serviços na Internet. O tráfego de voz foi modelado por fontes do tipo atividade-silêncio (On-Off). O serviço oferecido a tráfegos de voz pelos PHBs de Encaminhamento Expresso (EF) e Encaminhamento Assegurado (AF) foram avaliados. Também foi comparado o atendimento a estes tráfegos quando a implementação do serviço EF é realizado por diferentes mecanismos de escalonamento de filas, tais como PRR e WRR. A influência do tamanho dos pacotes do serviço de melhor esforço no atraso e *jitter* dos pacotes de voz atendidos pelo PHB AF foi discutida. Os resultados destacam a possibilidade de ganhos estatísticos da ordem de 70% para o atendimento conjunto de tráfegos de voz e melhor esforço. Como perspectivas para este trabalho estão a avaliação dos resultados para um maior número de fluxos, tanto TCP quanto fluxos de voz.

Referências

[1] Blake, S., Black, D. L., Carlson, M., Davies, E., Wang, Z., e Weiss, W. An architecture for differentiated services. *Internet RFC 2475* (dezembro de 1998).

[2] Braden, R., Clark, D. D., e Shenker, S. Integrated services in the Internet architecture: an overview. *Internet RFC 1633* (junho de 1994).

[3] Clark, D. D., e Fang, W. Explicit allocation of best effort packet delivery service. *IEEE/ACM Transactions on Networking 6*, 4 (agosto de 1998), 362–373.

[4] de Rezende, J. F. Avaliação do serviço assegurado para a diferenciação de serviços na Internet. In *Anais do XVII Simpósio Brasileiro de Redes de Computadores, SBRC'99* (Salvador, BA, maio de 1999), pp. 339–353.

[5] Floyd, S., e Jacobson, V. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking 1*, 4 (agosto de 1993), 397–413.

[6] Heinanen, J., Baker, F., Weiss, W., e Wroclawski, J. Assured forwarding PHB group. *Internet Draft* (fevereiro de 1999). `draft-ietf-diffserv-af-06`.

[7] Jacobson, V., Nichols, K., e Poduri, K. An expedited forwarding PHB. *Internet Draft* (fevereiro de 1999). `draft-ietf-diffserv-phb-ef-02`.

[8] Network Simulator (NS). `http://www-mash.cs.berkeley.edu/ns/`.

[9] Nichols, K., Blake, S., Baker, F., e Black, D. L. Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers. *Internet RFC 2474* (dezembro de 1998).

[10] Ziviani, A., de Rezende, J. F., e Duarte, O. C. M. B. Tráfego de voz em um ambiente de diferenciação de serviços na Internet. In *Anais do XVII Simpósio Brasileiro de Redes de Computadores, SBRC'99* (Salvador, BA, maio de 1999), pp. 354–367.